the CATO JOURNAL

AN INTERDISCIPLINARY JOURNAL OF PUBLIC POLICY ANALYSIS

THE POLITICAL ECONOMY OF POVERTY

James Gwartney and Thomas S. McCaleb Have Antipoverty Programs Increased Poverty?

Dwight R. Lee The Politics of Poverty and the Poverty of Politics

Karl Brunner The Poverty of Nations

Victor A. Canto Property Rights, Land Reform, and Economic Well-Being

David Osterfeld Resources, People, and the Neomalthusian Fallacy

Thomas Rustici A Public Choice View of the Minimum Wage

Norman P. Barry In Defense of the Invisible Hand

OTHER ARTICLES

James C. Miller III End the Postal Monopoly

Loren E. Lomasky Is Social Security Politically Untouchable?

Dale H. Gieringer The Safety and Efficacy of New Drug Approval

Charles W. Baird Labor Law and the First Amendment

Robert Cervero Deregulating Urban Transportation

Simon Rottenberg Job Protection in Urban Mass Transit

Leland B. Yeager Rights, Contract, and Utility in Policy Espousal

Gary M. Anderson and Robert D. Tollison Life in the Gulag: A Property Rights Perspective

Jack High State Education: Have Economists Made a Case?

Stuart Butler Privatization: A Strategy to Cut the Budget

VOLUME 5 NUMBER 1 • PUBLISHED BY THE CATO INSTITUTE • SPRING/SUMMER 1985

the CATO JOURNAL

VOLUME 5 NUMBER 1 • PUBLISHED BY THE CATO INSTITUTE • SPRING/SUMMER 1985

Editor
JAMES A. DORN

Assistant Editor
CATHERINE ENGLAND

Production
DAVID LAMPO

EDITORIAL BOARD

TERRY L. ANDERSON
Montana State University

CHARLES W. BAIRD
California State University
at Hayward

T. E. BORCHERDING
Claremont Graduate School

YALE BROZEN
University of Chicago

KARL BRUNNER
University of Rochester

JAMES M. BUCHANAN
Center for Study of
Public Choice
George Mason University

EDWIN DIAMOND
Massachusetts Institute
of Technology

ARTHUR A. EKIRCH
State University
of New York at Albany

RICHARD A. EPSTEIN
University of Chicago

STEVE H. HANKE
Johns Hopkins University

C. LOWELL HARRISS
Columbia University

FRIEDRICH A. HAYEK
University of Freiburg

LEONARD P. LIGGIO
Institute for
Humane Studies

SAM PELTZMAN
University of Chicago

ALVIN RABUSHKA
Hoover Institution
Stanford University

EARL C. RAVENAL
Georgetown University

VERNON L. SMITH
University of Arizona

ROLAND VAUBEL
Mannheim University

KAREN I. VAUGHN
George Mason University

RICHARD E. WAGNER
Florida State University

LAWRENCE H. WHITE
New York University

LELAND B. YEAGER
Auburn University

The CATO JOURNAL (ISSN 0273-3072) is published in the spring/summer, fall, and winter by the Cato Institute, 224 Second Street SE, Washington, D.C. 20003. Annual subscriptions are $18.00 for both individuals and institutions. Single copies are $6.00. Two-year subscriptions are $36.00 and three-year subscriptions are $54.00. Foreign subscribers should add $5.00 per year for regular delivery and $10.00 per year for airmail delivery.

Correspondence regarding subscriptions, changes of address, procuring back issues, advertising and marketing matters, etc., should be addressed to the Publications Department. All other correspondence, including requests to quote or reproduce material, should be addressed to the editor.

Unsolicited manuscripts cannot be returned and will be acknowledged only if accompanied by a stamped, self-addressed envelope. There is no submission fee. Articles should focus on specific public policy issues. Accepted manuscripts must conform to the CATO JOURNAL's style requirements. (A style sheet is available on request.)

Authors should submit three copies of their papers or communications to the editor. An abstract of no more than 250 words should also be included to aid referees. Manuscripts should have wide margins and be double-spaced throughout, including footnotes.

Unsolicited book reviews cannot be returned and will be acknowledged only if they are accepted for publication. They should conform to the above requirements, and should provide an in-depth analysis of major studies of public policy. Reviews should range from 800 to 1000 words.

The views expressed by the authors of the articles are their own and are not attributable to the editor, the editorial board, or the Cato Institute.

The CATO JOURNAL is regularly abstracted in *Current Contents/Social and Behavioral Sciences* and *Social Science Citation Index,* both produced by the Institute for Scientific Information. The CATO JOURNAL is also indexed and abstracted in the ABI/INFORM database. The CATO JOURNAL is available on microfilm from University Microfilms International, Ann Arbor, Michigan 48106.

Printed in the United States of America.

Copyright © 1985 by the Cato Institute
All rights reserved.

CATO JOURNAL VOLUME 5 NUMBER 1 SPRING/SUMMER 1985

THE POLITICAL ECONOMY OF POVERTY

James Gwartney and Thomas S. McCaleb
Have Antipoverty Programs Increased Poverty? 1

Dwight R. Lee
The Politics of Poverty and the Poverty of Politics 17

Karl Brunner
The Poverty of Nations 37

Victor A. Canto
Property Rights, Land Reform, and Economic Well-Being 51

David Osterfeld
Resources, People, and the Neomalthusian Fallacy 67

Thomas Rustici
A Public Choice View of the Minimum Wage 103

Norman P. Barry
In Defense of the Invisible Hand 133

OTHER ARTICLES

James C. Miller III
End the Postal Monopoly 149

Loren E. Lomasky
Is Social Security Politically Untouchable? 157

Dale H. Gieringer
The Safety and Efficacy of New Drug Approval 177

Charles W. Baird
Labor Law and the First Amendment 203

Robert Cervero
Deregulating Urban Transportation 219

Simon Rottenberg
Job Protection in Urban Mass Transit 239

Leland B. Yeager
Rights, Contract, and Utility in Policy Espousal 259

Gary M. Anderson and Robert D. Tollison
Life in the Gulag: A Property Rights Perspective 295

Jack High
State Education: Have Economists Made a Case? 305

Stuart Butler
Privatization: A Strategy to Cut the Budget 325

COMMUNICATIONS

Bruce R. Bolnick
The National Budget Referendum: Proceed with Caution 337

Reply by **Roger A. Arnold and Thomas L. Wyrick** 345

BOOK REVIEWS

Political Economy and Freedom
G. Warren Nutter
Reviewed by **Gertrude E. Schroeder** 351

Power and Privilege: Labor Unions in America
Morgan O. Reynolds
Reviewed by **Thomas J. DiLorenzo** 353

The Theory of Public Choice–II
Edited by James M. Buchanan and Robert D. Tollison
Reviewed by **Marilyn R. Flowers** 357

HAVE ANTIPOVERTY PROGRAMS INCREASED POVERTY?

James Gwartney and Thomas S. McCaleb

Introduction

For two decades following the end of World War II, rapid growth, rising incomes, and declining poverty rates characterized the American economy. Real per capita disposable income increased by 32.6 percent between 1950 and 1965. Mean family income adjusted for inflation was 54 percent higher in 1965 than in 1950. Furthermore, the rapid growth and rising incomes were not confined to the nation's middle and upper income families. In fact, the largest income gains were registered by the bottom fifth of income recipients. Measured in constant 1972 dollars, the mean income of the bottom one-fifth of families rose from $1411 in 1950 to $2506 in 1965, an increase of almost 78 percent (U.S. Department of Commerce 1983, Table 17).

Steady progress was made against poverty during the period. By 1965 only 13.9 percent of American families were officially classified as poor, down from 32 percent in 1947 and 18.5 percent in 1959. Thus, in less than a generation economic progress had cut the overall poverty rate in half. Moreover, as Table 1 illustrates, the poverty rate declined steadily in all age groups. Among the elderly, the poverty rate fell from 57 percent in 1947 to 22.8 percent in 1965. For youthful families, those headed by an individual under age 25, the poverty rate fell from 45 percent in 1947 to 26.9 percent in 1959 and to 19.4 percent in 1965. The pattern for families in the prime working-age groups is similar.

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

James Gwartney is Professor of Economics and Policy Sciences and Thomas S. McCaleb is Associate Professor of Economics at Florida State University. Mr. McCaleb is also a former Senior Staff Economist with the President's Council of Economic Advisers. This research was partially supported by the Political Economy Research Center.

TABLE 1

THE OFFICIAL POVERTY RATE OF FAMILIES BY AGE OF HOUSEHOLDER FOR SELECTED YEARS, 1947–1982

Age of Householder	1947	1959	1965	1968	1970	1975	1980	1982
Under 25	45.0	26.9	19.4	13.2	15.5	21.0	21.8	26.1
25–44	27.0	16.5	12.8	9.3	9.5	10.3	11.8	14.2
45–54	27.0	15.0	9.6	7.0	6.6	6.6	7.6	8.9
65 & over	57.0	30.0	22.8	17.0	16.5	8.9	9.1	9.3
All Families	32.0	18.5	13.9	10.0	10.1	9.7	10.3	12.2

SOURCE: The 1947 data are from *Economic Report of the President: 1964*, Table 7. All other data are from U.S. Department of Commerce: Bureau of the Census, *Estimates of Poverty Including The Value of Noncash Benefits: 1979 to 1982.*

This progress, achieved with little assistance from income transfer programs and government antipoverty efforts, was thought to be insufficiently rapid, however. In 1964 President Lyndon Johnson called for a "Marshal Plan" commitment by the government to eradicate poverty within the next generation. He stated, "We cannot and need not wait for the gradual growth of the economy to lift this forgotten fifth of our Nation above the poverty line" (Council of Economic Advisers 1964, p. 15).

In the mid-1960s, poverty was widely perceived as a problem of inadequate opportunity and bad luck. Political entrepreneurs believed that it could be solved by taxing the nonpoor and using the revenues to provide education, job training, health care, and income maintenance for the poor. The very prosperity and economic growth that produced the dramatic decline in the incidence of poverty prior to 1965 also increased the attractiveness of government transfer programs as a device for bringing forth yet further reductions in poverty. When almost everyone is poor, taxing the rich to improve the welfare of the poor has only a negligible effect. There simply are not enough rich to do much about poverty. Increasing income, however, makes it appear feasible to do something about the status of low-income individuals and families. The 1964 *Economic Report of the President* (p. 77) expressed the dominant view:

> Conquest of poverty is well within our power. About $11 billion a year would bring all poor families up to the $3000 income level we have taken to be the minimum for a decent life. The majority of the nation could simply tax themselves enough to provide the necessary income supplements to their less fortunate citizens. The burden—

one fifth of the annual defense budget, less than 2 percent of GNP—would certainly not be intolerable.

Growth of the Transfer Society

As the War on Poverty programs began in the latter half of the 1960s, there was a virtual explosion in government transfers. In 1965, transfers accounted for 7.5 percent of personal income, up only slightly from 6.7 percent in 1950. By 1975 total cash transfer payments had soared to 14.1 percent of personal income (see Table 2). Measured in inflation-adjusted dollars, cash transfers nearly tripled in just 10 years, expanding from $105.5 billion in 1965 to $273 billion in 1975.

Even these numbers do not reflect the full extent of the increased government involvement in fighting poverty. As Table 2 illustrates, the growth of means-tested assistance targeted toward the poor was even more rapid as programs providing in-kind benefits such as food, health care, and housing for low-income households were established and expanded. Measured in 1980 dollars, expenditures on these means-tested, noncash benefits rose from $4.6 billion in 1965 to $30.3 billion in 1975, an increase in real terms of 559 percent in

TABLE 2

GROWTH OF TRANSFER PAYMENTS, 1965–1982
(Billions of Constant 1980 Dollars)

Type of Benefit	1965	1970	1975	1980	1982
Total Cash Transfer Payments[a]	105.5 (7.5)	170.0 (9.9)	273.0 (14.1)	297.2 (13.8)	319.7 (14.6)
Means-tested Cash Assistance[b]	14.7	15.1	27.4	23.6	22.1
Noncash, means-tested	4.6	16.8	30.3	39.1	40.0
Food stamps	0.1	1.2	6.7	8.7	8.7
School lunches	NA	0.3	1.3	1.5	1.5
Public housing	0.9	3.5	3.5	4.5	4.3
Medicaid	3.6	11.9	18.8	24.3	25.5
Total, means-tested	19.3	31.9	57.7	62.7	62.2

[a] Numbers in parentheses represent total cash transfer payments as a percent of personal income.

[b] Includes Aid to Families with Dependent Children, General Assistance, Supplemental Security Income, and means-tested veteran's pensions.

SOURCE: U.S. Department of Commerce: Bureau of the Census, *Estimates of Poverty Including the Value of Noncash Benefits: 1979–1982*, Table A, and *Economic Report of the President, 1983*, Table B-22.

just 10 years. While expenditures on means-tested cash assistance actually declined during the 1975–1982 period, this decline was more than offset by continued growth, albeit at a slower rate, in noncash benefits. By the early 1980s, total means-tested expenditures in *real* terms were nearly three and one-half times the 1965 level.

Poverty Rates and the War on Poverty

Did the increase in expenditures and expansion of government antipoverty programs accelerate the decline in the poverty rate as President Johnson had anticipated? Incredible as it may seem, almost the opposite occurred. Just as government spending on the various antipoverty programs accelerated in the late 1960s, progress against poverty came to a grinding halt. The official poverty rate reached a minimum in the late 1960s. By 1980, the overall rate was 10.3 percent, virtually unchanged from the 10.0 percent rate of 1968. Moreover, as is clear from Table 1, the aggregate rate conceals important differences across age groups that have heretofore gone largely unnoticed. During the 1970s, the official poverty rate for the elderly continued to decline, from 17.0 percent in 1968 to 9.1 percent in 1980. The rate for families headed by an individual aged 45–54 increased marginally, from 7.0 percent to 7.3 percent. In contrast, the poverty rate for the other age groups increased significantly. By 1980 the official poverty rate for families in the 15–24 age grouping had risen to 21.8 percent (821,000 households), up from 13.2 percent in 1968. Similarly, the incidence of poverty among families headed by persons age 25–44 rose from 9.3 percent in 1968 to 11.8 percent in 1980 (3,168,000 households). Families headed by a person under age 45 now account for nearly two-thirds of the poor families in the United States. The increases in the official poverty rates for working-age households have continued into the 1980s. Thus, during the War on Poverty era, consistent progress against poverty has been limited to the elderly.

The official poverty rate reflects the percentage of families with cash income below a certain threshold amount. This poverty income level has been adjusted over time for inflation.[1] The official poverty rate does not take into account noncash (in-kind) transfer payments. Because the noncash benefits have grown so much more rapidly than cash transfers since the inception of the War on Poverty, some analysts have argued that the official rate is a misleading indicator of changes in the number of families living in poverty.[2] A recent study by the U.S. Department of Commerce (1984) provides data for 1979–1982 on the poverty rate adjusted for noncash benefits and on noncash expenditures adjusted for inflation for the period 1965–1982. Given the size of the noncash transfers and the impact of the in-kind benefits on the adjusted poverty rates for 1979–1982, an adjusted rate for each group can be reconstructed for earlier years.[3]

---

[1] The Social Security Administration (SSA) developed the official definition of poverty. Since consumption survey data indicated that low and median income families of three or more persons spent approximately one-third of their income on food, the SSA established the official poverty income level at three times the cost of an economical, nutritionally adequate food plan for a family of a specified size. This method was used to derive the poverty income threshold adjusted for factors such as family size, sex of the family head, number of children under 18 years old, and farm or nonfarm residence. The poverty threshold is adjusted annually to account for rising prices. The chart below

The adjustments for noncash benefits are made in two alternative ways. One values in-kind transfers at *market value*, that is, what it costs to supply the goods and services to the recipient. The other attempts to estimate the *recipient value* or the amount of cash income for which recipients would be willing to trade their right to the noncash benefits. Because market value is relatively easy to compute, it is most often used in studies seeking to value in-kind benefits. However, since the recipient value method values noncash benefits at their estimated cash-equivalent to the recipients, it is accepted by most analysts as the more appropriate measure.

Adjusted poverty rates using both methods are presented in Table 3. Adjustment lowers the rate during the 1970s but the adjusted rate nevertheless follows a time path that is quite similar to that of

---

illustrates how the poverty income threshold for a family of four persons has increased as prices have risen:

Year	Average Poverty-Income Threshold for a Family of Four
1959	$2973
1965	3223
1970	3968
1975	5500
1980	8414
1982	9862

[2] See Paglin (1979) and Browning (1975) for a presentation of this view.

[3] Mathematically, the adjusted poverty rate within age group a for each year i is equal to:

$$APR_{ai} = OPR_{ai} - \frac{MTE_i}{MTE_{80}}(OPR_{a80} - APR_{a80})$$

where APR_{ai} = adjusted poverty rate during year i within age group a
OPR_{ai} = official poverty rate during year i within age group a
MTE_i = noncash, means-tested expenditures in year i
MTE_{80} = noncash, means-tested expenditures in 1980
OPR_{a80} = official poverty rate within age group a in 1980
APR_{a80} = adjusted poverty rate within age group a in 1980.

TABLE 3

THE POVERTY RATE OF FAMILIES ADJUSTED FOR NONCASH BENEFITS BY AGE OF HOUSEHOLDER FOR SELECTED YEARS, 1959–1982

Age of Householder	1959[a]	1965	1968	1970	1975	1980	1981	1982
Adjusted By Market Value Method[b]								
Under 25	26.9	18.5	10.9	12.4	15.3	14.5	16.6	18.7
25–44	16.5	12.3	7.9	7.6	6.9	7.4	8.6	10.1
45–54	15.0	9.0	6.3	5.7	4.9	5.1	6.2	6.2
65 & over	30.0	22.0	15.0	13.7	3.8	2.5	2.5	2.6
All Families	18.5	13.4	8.7	8.3	6.4	6.1	7.2	8.0
Adjusted By Recipient Value Method[b]								
Under 25	26.9	19.0	12.3	14.2	18.7	18.8	20.9	24.0
25–44	16.5	12.5	8.6	8.5	8.5	9.5	10.7	12.3
45–54	15.0	9.5	6.7	6.1	5.8	6.2	7.5	8.0
65 & over	30.0	22.4	15.9	14.9	6.0	5.4	5.1	5.5
All Families	18.5	13.7	9.3	9.2	8.1	8.2	9.2	10.2

[a] The poverty rate for 1959 is the official rate taken from Table 1 of the text.
[b] See text, footnote 3, for a description of how the adjusted rates were derived for 1965, 1968, 1970, and 1975.

SOURCE: U.S. Department of Commerce, Bureau of Census, *Estimates of Poverty Including The Value of Noncash Benefits: 1979 to 1982.*

the official rate. For the elderly, the adjusted rate fell throughout the period, reaching 2.1 percent by 1980 using the market value method and 5.4 percent using recipient value.

For families in the under-25 age group, however, the adjusted poverty rate in 1980 was well above the 1968 level using both methods. In fact, the recipient value poverty rate in 1980, 18.8 percent, was almost exactly equal to the 19.0 percent rate in 1965. Thus, 15 years after the explosion in government transfer payments began, the poverty rate for youthful families was essentially unchanged from its 1965 level even after allowance is made for the in-kind benefits. The rise in the adjusted poverty rate for this age group continued into the 1980s, reaching 24.0 percent in 1982.

For the 25–44 age group, the market-value adjusted rate stood at 7.4 percent in 1980, down only slightly from 7.9 percent in 1968. Using the recipient value method, however, the adjusted rate in 1980 was 9.5 percent, well in excess of the 8.6 percent rate in 1968. While the adjusted poverty rate for the 45–54 age grouping did not begin climbing until the latter half of the 1970s, by the early 1980s it too had risen above the levels of the late 1960s.

Thus, whether one looks at the official or adjusted poverty rates, the picture is the same. Except for the elderly, soon after the massive increase in transfer payments in the late 1960s, the steady progress of the pre-War on Poverty era came to a halt and the poverty rates of working age Americans began to rise.

Why Have Poverty Rates Risen?

Can the increasing poverty rates be explained by economic conditions? Per capita real GNP grew at an annual rate of 2.1 percent during the 1970s, a modest reduction from the 2.6 percent rate of the 1960s. However, the decade of slowest growth during the post-World War II era was the 1950s, when per capita real GNP rose by only 1.5 percent per year. Despite the slower growth during the 1950s, the poverty rate fell among all age groups. In addition, when the 1950–1980 period is divided into equal subperiods, the average annual rates of growth in per capita real GNP during the two subperiods are identical, 2.1 percent.[4]

Furthermore, financing the expanding transfer sector during the War on Poverty era required higher taxes which impede the efficient use of resources, reduce productivity, and retard economic growth. Thus, to the extent that the growth rate in the U.S. slowed during the 1970s, the expanding tax-transfer sector itself was at least partially responsible.[5]

If the state of the economy fails to explain the lack of progress against poverty in the face of massive increases in government transfer payments, what does? The application of simple economic theory suggests that the seeds of failure lie within the War on Poverty programs themselves. While government transfer programs have improved the standard of living for some of the poor through increased benefit levels and relaxed eligibility requirements, they have also stifled the incentives for the poor to improve their own economic status and for the nonpoor to avoid poverty. They have introduced a perverse incentive structure, one that penalizes self-improvement and protects individuals against the consequences of their own bad choices.

---

[4] Per capita *real* disposable personal income actually grew more rapidly during the latter period. It grew at an annual rate of 2.4 percent during the 1965–1980 period, up from 1.9 percent during the 1950–1965 era.

[5] See Lee (1985) for evidence on the impact of rising tax rates on economic growth during the 1970s.

Recent studies of the Aid to Families with Dependent Children (AFDC) program using panel data show that there are two quite distinct groups of poor. One group, whom we shall call the hardcore poor, remain on welfare for many years. The second group, whom we shall call the marginal poor, participate in welfare programs for much shorter periods of time, often exiting the programs in less than one year. Indeed, only about half of all AFDC spells last beyond one year, with 16–18 percent surviving to the fifth year, and 5–7 percent to the tenth year (O'Neill 1984, p. 6).[6]

The hardcore poor are often victims of personal misfortune; their poverty is usually attributable to such factors as debilitating injury, disease, or physical, mental, or emotional disability. By and large, their poverty is not the result of their own decisions and actions, and their ability to improve their economic status is quite limited. In contrast, the marginal poor move into and out of poverty. Their poverty is sometimes due to temporary impediments to work such as injury, disease, economic recession, or a change in family status. In other cases, their poverty results from inappropriate individual choices or errors of judgment such as pregnancy or premature termination of schooling. Unlike the hardcore poor, the marginal poor retain some control over their economic status and they respond to economic incentives. Improving job opportunities reduce the incidence of poverty among the marginal poor, for example, while the hardcore poor remain poor through recessions and recoveries alike.

The structure of the War on Poverty transfer programs produces four effects, each of which operates to retard progress against poverty and increase the incidence of poverty among the marginal poor.

The Higher Real Benefit Effect

For individuals with few skills and low wages, antipoverty benefits offer an alternative to income from employment. Increases in the real value of benefit payments make dependency on the government even more attractive compared with the alternative of self-support. Thus, while the higher benefits tend to reduce the severity of poverty, they also increase its incidence. Just as unemployment compensation raises the unemployment rate by increasing the duration of unemployment,[7] higher antipoverty benefits raise the poverty rate by increasing the duration of poverty among the marginal poor (O'Neill 1984).

---

[6] See Bane and Ellwood (1983a, 1983b) for complementary findings.

[7] See Ehrenberg and Oaxaca (1976), Feldstein (1978), and Classen (1977).

The expansion in total spending on transfer programs during the War on Poverty era not only increased real benefits per recipient, but as Table 4 shows, the benefit levels for many transfer programs also rose more rapidly than per capita disposable personal income. For example, while real personal income rose by 41 percent during the 1965–1980 period, the real value of per recipient food stamp benefits increased by 136 percent. Measured in 1980 dollars, the average annual Medicare hospitalization payment rose by 207 percent over the same period, and medical benefits increased by 361 percent between 1967 (the initial year for these programs) and 1980. Social Security benefits also grew more rapidly than personal income—53.6 percent in real terms compared to 41.0 percent—during the 1965–1980 period.

Although payments made to participants in means-tested cash transfer programs such as AFDC and Supplementary Security Income (SSI) grew more slowly than personal incomes, benefits did rise faster than the rate of inflation so that the purchasing power of the cash benefits nevertheless increased. In addition, by the 1970s recipients of AFDC and SSI in most states were also eligible for in-kind benefits from food stamps, Medicaid, housing programs, and school lunch subsidies. Once the in-kind benefits are taken into account, the real transfer incomes of many individuals participating in the AFDC and SSI programs also expanded more rapidly than per capita disposable personal income. These higher benefit levels induce some individuals to substitute income transfers for employment earnings, resulting in a greater incidence of poverty and a higher poverty rate.

TABLE 4

GROWTH OF PER RECIPIENT REAL BENEFIT LEVELS, 1965–1980

Program	Benefits Per Recipient[a] 1965	1980	Increase (%)
Food Stamps	201	475	136.6
Medicare (hospital)	319[b]	981	207.4
Medicare (medical)	93[b]	427	361.2
Social Security (monthly)	221	339	53.6
Personal Income (per capita)	6246	8808	41.0

[a] Benefits per recipient are expressed in constant 1980 dollars.
[b] Data are for 1967, the initial year of the program.

SOURCE: *Statistical Abstract of the United States* (Washington, D.C.: Government Printing Office, respective years).

The High Implicit Tax Effect

While the higher real benefits paid by the War on Poverty transfer programs make poverty more attractive, the programs have also reduced the incentives of the marginal poor to help themselves once they enter poverty. Indeed, these government programs penalize efforts at self-improvement because the benefits payable under the transfer programs aimed most directly at the poor decrease as earned income rises. As an individual or family increases employment earnings, benefits from programs such AFDC, food stamps, Medicaid, school lunches, and rent supplements fall.

The rate at which the benefit payments from each program are reduced per dollar of nontransfer income received is the implicit marginal tax rate associated with that program. For example, food stamp benefits are reduced by $30 for each $100 of monthly earnings up to $800. Families are eligible for free school lunches until income reaches 125 percent of the official poverty level, and reduced-price lunches are available until income reaches 195 percent of the official poverty level. The implicit marginal tax rates associated with cash transfer programs such as AFDC and unemployment compensation are high, typically in the 50 to 60 percent range. The marginal tax rates associated with individual programs, however, understate the disincentives faced by the recipients. Most of the poor receive benefits from several programs simultaneously. When the implicit marginal tax rates associated with each of these programs are considered together, the *compound multi-program implicit marginal tax rate* is very high.

For example, as Table 5 shows, a mother with two children and no earned income residing in Pennsylvania would qualify for annual cash and in-kind benefits of $7568 from AFDC, food stamps, Medicaid, and the Earned Income Tax Credit. If the family's earnings rose to $2000, transfer benefits would be reduced and taxes increased, leaving the family with spendable income of $8391. Thus, additional earned income of $2000 generates only $823 in additional spendable income, equivalent to a marginal tax rate of 58.8 percent. At higher levels of earned income, this implicit marginal tax rate is even greater. If earned income rose from $4000 to $5000, spendable income would *decrease* from $9214 to $7694, an implicit marginal tax rate of 252 percent. If income increased further to $6000, the family would lose eligibility for AFDC, and its after-tax-and-transfer income would decrease again, this time by $246. In fact, a family earning $6000 has less spendable income than a family with no earned income, and a

TABLE 5

THE EFFECT OF TRANSFER BENEFITS AND TAXES ON THE INCENTIVE OF A PENNSYLVANIA MOTHER WITH TWO CHILDREN TO EARN INCOME (SEPTEMBER 1983)

Annual Gross Wage	Transfer Benefits[a]	Income and Employment Taxes[b]	Spendable Income	Implicit Marginal Tax Rate
0	7568	0	7568	—
2000	6525	134	8391	58.8
4000	5482	268	9214	58.8
5000	3040	346	7694	252.0
6000	2059	611	7448	124.6
7000	1719	810	7909	53.9
8000	1378	1021	8357	55.2
9000	1038	1240	8798	55.9
10000	698	1469	9229	56.9

[a] The following benefits are included: AFDC, Earned Income Tax Credit, food stamps and Medicaid. The Medicaid benefits were valued at the 1978 national average adjusted for inflation between 1978 and 1982.

[b] Includes social security and federal and state income taxes.

SOURCE: Data are derived from U.S. House of Representatives: Committee on Ways and Means, *Background Material on Poverty* (Washington, D.C.: Government Printing Office, 1983) Table 10, page 89.

family earning $10,000 each year, equivalent to a full-time year-round job paying $5 an hour, would have spendable income of $9229, just $1661 more than a family with no earned income at all. The loss in transfer benefits and the increased taxes when earnings rise from zero to $10,000 is equivalent to a tax rate of 83 percent on earned income.

Such high implicit marginal tax rates pose a significant disincentive to work for those individuals whose potential earnings are relatively low.[8] Many transfer recipients who would otherwise engage in market work will decide to work fewer hours or not at all. Clearly, then, some portion of transfer income is merely replacement income—that is, it replaces income that the recipient would have earned in the absence of the high implicit marginal tax rate effect. The larger the size of this replacement component, the smaller the *net* positive impact of income transfers on the measured income of the poor.

---

[8] Marginal tax rates in the 40, 50, and 60 percent range have been shown to reduce the incentives of high income recipients to earn taxable income (Lee 1985). Paradoxically, many transfer recipients face implicit marginal rates in excess of those faced by taxpayers in the highest marginal tax brackets.

The Skill-Depreciation Effect

Individuals who have not utilized their skills for extended periods of time find it difficult to compete with otherwise similar individuals who have continuous labor force participation. Thus, the long-term consequences of the current programs will be even more destructive than the short-term effects. As the marginal poor spend more time dependent upon transfers and less time in the work force, their skills depreciate and their work record deteriorates. They become less and less able to support themselves. Over time, more and more of the marginal poor will be transformed into hardcore poor.

The data on labor force participation of householders living in poverty illustrate the importance of the skill depreciation effect. Between 1966 and 1980, the proportion of poor households headed by a person *who did not work at all during the year* rose from 39.7 percent to 49.6 percent. For female headed households in poverty, 61.5 percent of the household heads did not work at all in 1980, up from 52.7 percent in 1966. Thus, while the labor force participation of all females was climbing, the rate for poor females was declining.[9] Clearly, these ominous trends provide little reason for optimism that the poor will be better able to provide for themselves in the years ahead.

The rising poverty rate among youthful families (see Table 3) that has accompanied the growth of War on Poverty expenditures is particularly discomforting. As transfers make dependency more attractive relative to work experience, schooling, and other forms of human capital investment, youthful recipients fail to develop skills that have in the past enabled the young to escape from poverty. Withdrawal from the labor force by recipients in their teens and twenties will only increase the poverty rate among these individuals as they move into their thirties and forties. Clearly, the long-run destructive effects of the current system have not yet been fully realized.

The Moral Hazard Effect

The poverty of the marginal poor, unlike that of the hardcore poor, is at least in part the result of conditions over which they exercise some control. Some choose a lifestyle that increases the likelihood of poverty. Some consciously choose leisure over work; others are poor because they failed to complete school, or because of drug or alcohol abuse, or because they have chosen to bear children without the means to provide adequate support. These individuals have made choices that severely limit their ability to be independent and self-supporting.

---

[9] See Murray (1984) for additional evidence on the negative impact of transfer programs on the labor force participation of low-income families.

Unfortunately, it is not always easy to distinguish between the unfortunate and the irresponsible. At best, government programs make such distinctions in the grossest of terms. While income transfers increase the standard of living of the hardcore poor who are unable to help themselves, they also reduce the cost to others of making those very choices that are most likely to bring on poverty. In short, by offering "insurance" against adversity over which the marginal poor have some control, the programs encourage the very situations that they were designed to combat. In the insurance industry, this is referred to as the "moral hazard problem." Recognizing this problem, private insurance companies seldom offer protection against adversities the occurrence of which is substantially affected by the behavior of potential policyholders.[10]

Existing government transfer programs attempt to reduce the severity of the moral hazard problem by adopting detailed rules and guidelines intended to limit abuse and to promote legislative intent. The regulations, however, eliminate the possibility of solutions tailored to each individual recipient which would in fact minimize moral hazard. Indeed, the regulations encourage potential recipients to alter their behavior so as to meet program eligibility requirements rather than to make choices that would reduce or eliminate their need for income transfers.

The impact of the moral hazard effect manifests itself in several ways. For example, the number of "unrelated individuals" in poverty fell from 4.9 million in 1959 to 4.7 million in 1968 in spite of increasing population. Since 1968, however, the number has risen steadily, reaching 6.2 million in 1980, as dependence on government has replaced dependence on family. The rise in the poverty rate among families headed by persons in the 15–24 age group between 1965 and 1980 reflects the availability of government support for those who choose high-risk life styles (involving, for example, sexual promiscuity or drug or alcohol use) that families often refuse to subsidize. The availability of government transfers has also made it less costly for husbands in low-income households to desert and for unmarried fathers to avoid responsibility for their children. Thus, the transfer programs have contributed to the dramatic rise in the number of female-headed households in recent decades.[11] Interestingly, there is one eligibility requirement for government transfer benefits that cannot be satisfied by a change in an individual's behavior: age. Therefore, moral hazard is not a problem in programs for the elderly. This helps explain why, alone among age groups, the poverty rate for those over 65 has continued to decline during the War on Poverty era.

---

[10] For example, fire insurance policies do not cover damage due to a fire set by the policyholder. Similarly, the authors are unaware of any private company offering insurance against income loss due to dropping out of school, nonmedical absenteeism, alcoholism, or drug addiction, all of which are subject to the moral hazard effect.

Conclusion: An Agenda for Welfare Reform

While the intentions of the architects of the War on Poverty were noble, their approach to reducing the incidence of poverty has proven counterproductive. The current system of income transfers confronts the poor with perverse incentives that discourage self-help efforts in the short run and induces recipients to make decisions that retard their ability to escape poverty in the long run. The system unwittingly encourages behavior that leads to and perpetuates poverty. The need for policy reform is clear.

To be effective, public policy to alleviate poverty must be consistent with four basic principles. These four principles provide an agenda for welfare reform.

1. *Except for the handicapped and the elderly, all adults must be required to work in order to be eligible for long-term (say, more than 3 months) welfare benefits.* The concept of providing long-term public assistance to those who do not work was developed during a different environment. For men, it arose during the massive unemployment of the Great Depression; for women, it originated prior to the movement of large numbers of married women, even those with small children, into the labor market. A work requirement will substantially alleviate the negative long-term impact on the poor of both the skill-depreciation effect and the moral hazard effect.

2. *The welfare system must recognize the importance of the family, church, private charity, and community action in the alleviation of poverty.* In contrast with the government, private voluntary donors have the capacity to structure help for the poor in ways that minimize moral hazard and avoid the perverse incentive effects. Private action

---

[11] For whites, the proportion of families headed by a female rose from 8.8 percent in 1967 to 11.9 percent in 1981. For blacks, the proportion of female-headed households rose from 23.6 percent in 1967 to 41.7 percent in 1981. While the trend has been upward since World War II, the increase has been much more rapid since the mid-1960s.

complements economic growth in reducing poverty; government transfer programs impede the capacity of economic growth to reduce the incidence of poverty by retarding growth itself. Apart from temporary assistance such as short-term unemployment compensation, government transfer programs should be directed only at the hardcore poor—handicapped adults and disadvantaged children—for whom perverse incentives and moral hazard are much less relevant.

3. *The welfare system should be structured to reenforce certain traditional values that encourage individual, parental, and family responsibility.* Throughout history, certain social values have played an important role in meeting the needs that accompany economic hardship. It is indeed appropriate for a free society to seek to establish institutional arrangements that will make individuals accountable for their actions. The idea that society should pay for individual irresponsibility is counterproductive. It has undermined progress against poverty in the past and, unless rejected, it will continue to do so in the future.

4. *Transfer recipients should not be allowed to use children as hostages in order to blackmail society.* No disadvantaged child is undeserving. However, we must find ways of helping children that will not be used and abused by undeserving, irresponsible adults. Transfer programs targeted directly at children and a work requirement associated with receipt of AFDC would reduce the misuse of children by adults seeking income transfers for themselves.

The problem of poverty continues to fester not because we are failing to do enough, but rather because we are doing so much that is counterproductive. We must not continue to cast the issue in terms of the compassion of tax-transfer proponents versus the callousness of their critics. If we want to do something *effective*, the issue must be redrawn between those who would continue the government programs of the past and those who seek new approaches with long-lasting benefits and fewer harmful side-effects. The current approach treats the symptoms, but it also unwittingly promotes the disease. In contrast, while still providing for the welfare of those who truly cannot help themselves, the principles we have enunciated would encourage self-help and discourage long-term dependency and personal irresponsibility. The future welfare of the poor and the vitality of our economy may well depend on how we resolve this important issue.

References

Bane, Mary Jo, and Ellwood, David T. *The Dynamics of Dependence: The Routes to Self-Sufficiency.* Report prepared for U.S. Department of Health and Human Services, June 1983a.

Bane, Mary Jo, and Ellwood, David T. "Slipping into and out of Poverty: The Dynamics of Spells." National Bureau of Economic Research working paper, September 1983b.

Browning, Edgar K. *Redistribution and the Welfare System.* Washington, D.C.: American Enterprise Institute, 1975.

Classen, K. P. "The Effect of Unemployment Insurance on the Duration of Unemployment and Subsequent Earnings." *Industrial and Labor Relations Review* 30 (July 1977): 438–44.

Council of Economic Advisers. *Economic Report of the President: 1964.* Washington, D.C.: Government Printing Office, 1964.

Ehrenberg, R. G., and Oaxaca, R. L. "Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Gain." *American Economic Review* 66 (December 1976): 754–66.

Feldstein, Martin. "The Effect of Unemployment Insurance on Temporary Layoff Unemployment." *American Economic Review* 68 (December 1978): 834–46.

Lee, Dwight, ed. *The Political Economy of Capital Formation.* San Francisco: Pacific Institute, 1985.

Murray, Charles. *Losing Ground: American Social Policy, 1950–1980.* New York: Basic Books, 1984.

O'Neill, June A. et. al. *Analysis of Time on Welfare.* Washington, D.C.: Urban Institute, 1984.

Paglin, Morton. *Poverty and Transfers In-Kind.* Palo Alto: Hoover Institution Press, 1979.

U.S. Department of Commerce: Bureau of the Census. *Money Income of Households, Families, and Persons in the United States: 1981.* Washington, D.C.: Government Printing Office, 1983.

U.S. Department of Commerce: Bureau of the Census. *Estimates of Poverty Including the Value of Noncash Benefits: 1979 to 1982.* Washington, D.C.: Government Printing Office, 1984.

U.S. House of Representatives: Committee on Ways and Means. *Background Material on Poverty.* Washington, D.C.: Government Printing Office, 1983.

THE POLITICS OF POVERTY AND THE POVERTY OF POLITICS

Dwight R. Lee

> [T]he poor [may] deserve more, but if the government tries to provide more, it will not do anyone any good.
>
> —Nathan Glazer[1]

Introduction

The notion that we have to depend on government to assist the poor has acquired the status of revealed truth. Even those who acknowledge the unparalleled success of market economies at creating wealth are uneasy about, if not outright hostile to, the market distribution of this wealth. There can be no denying that some people will be left behind by market competition. Indeed, the very success of the market at creating wealth comes from the fact that it constantly threatens people with poverty; and when consumers signal with their thumbs down, the threat is carried out without mercy.

But even if it is assumed that the market fails to generate a distribution of income that most people find satisfactory, does this justify government programs to promote a more acceptable income distribution? The answer depends crucially on whether or not there are reasons for believing that such government programs will improve matters. If, for example, government welfare programs impose a heavy burden on economic productivity, yet are incapable of changing the distribution of income in a more acceptable direction, then the "failure" of the market with regards to income distribution would provide no justification for government intervention in the market process.

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Associate Professor of Economics at George Mason University and a Research Fellow at the Center for Study of Public Choice.

[1] "Reagan's Social Policy—A Review," *The Public Interest* (Spring 1984): 94.

The importance of comparing realistically the income distribution consequences of the market process with those of the political process may seem obvious. But in fact, such a comparison is seldom made. The market process is seen to be driven by competition between self-seeking individuals who are unconcerned about the impact of their decisions on the overall distribution of income. Income inequality is correctly judged to be endemic to the market process. On the other hand, the political process is seen to be driven by concern for broad social objectives. The prime motivation for government transfer programs is assumed to be to help the poor and reduce income inequality. If these programs fail in what they were set up to accomplish, the problem is seldom seen to be anything inherent in the political process that spawned the programs.

The evidence is strong that government transfer programs have failed. The U.S. Bureau of the Census announced in August, 1984 that 35.3 million people, or 15.2 percent of the population, were living in poverty. This is a startling statistic given the enormous amounts the government is spending on social welfare programs. Though it is difficult to pin down the total welfare bill with precision, it has been estimated that it came to over $403 billion in 1982.[2] This amounts to $11,730 per officially designated poor person, or $46,920 for each poor family of four.

But predictably, the blame for this blatant failure seldom cuts to the fundamentals of the problem: a realistic assessment of the political process. At the partisan level, the blame is placed on insufficient funds for these programs, with the Reagan budget "cut backs" being singled out for special criticism. Although funding was reduced for some poverty programs during the early years of the Reagan administration, overall spending for poverty programs continued upward during the 1980s. More important, since 1973 there has been a persistent increase in the number of poor as measured by the Bureau of the Census. An explanation of the existing level of poverty, given the amounts being spent to eliminate it, requires something more sophisticated than an attack on the Reagan administration for being insensitive. The seriousness with which this attack is made, and widely accepted as justified, reflects the reluctance to subject the political process behind transfer programs to the same critical examination to which the market process is subjected.

---

[2] This figure comes from Jonathan Hobbs, "Welfare Need and Welfare Spending," Heritage Foundation Backgrounder, Washington, D.C., 13 October 1982, and includes only the welfare component of Social Security payments.

Even those most critical of the government welfare programs seldom see the problems as inextricably tied to the political process. Favorite targets are fraud and corruption that should be rooted out with tighter controls over existing programs. Others see the solution coming from reforming existing programs; such as reducing reliance on in-kind transfers with more aid being provided in the form of cash payments.[3] But such calls for reform have been made for decades, and they have always been rendered politically impotent. In any event, welfare fraud can account for only a small amount of the cost of our welfare industry. And given the outpouring of scholarly articles on the poverty question, it is difficult to argue that the failures of our poverty programs can be reversed with yet more advice on desirable reforms.

It is not, then, the purpose of this paper to suggest welfare reform. Rather, an effort will be made to explain why poverty has been, and will continue to be, amazingly immune to political measures to alleviate it. In order to accomplish this, we will subject the political process to the same type of analytical scrutiny which economists have always applied to the market process. This means incorporating the same motivational assumption into our political analysis as is incorporated into standard economic analysis: namely, that individuals are motivated predominantly, though not solely, by considerations of self-interest.[4] Proceeding with this self-interest assumption, we will be led to the conclusion that political activity is not likely to significantly alter the distribution of income away from that which results from market activity. Furthermore, this theoretically arrived at conclusion is shown to be empirically supported.

The implication of this is rather depressing. If government transfer programs reduce the productivity of the economy (which they obviously do), and if they fail to alter the distribution of income in favor of the poor, it necessarily follows that these programs have made the poor worse off. Even more depressing is the inability to respond appropriately to this unfortunate situation through the means of ordinary politics. As will be seen, it is the very failure of social welfare programs that explains their political viability.

---

[3] See Hobbs, "Welfare Need." Also for a list of suggested reforms in our approach to social welfare spending, see John C. Goodman, "Poverty and Welfare" in *To Promote Prosperity: U.S. Domestic Policy in the Mid-1980s*, ed. John Moore (Stanford, Calif.: Hoover Institution Press, 1984).

[4] Even if one chooses to quibble with this assumption, no one can argue persuasively that objectivity is served by applying it to the market process but not to the political process.

Who Benefits from Poverty Programs?

It is useful to begin by asking: What is the motivating political force behind government poverty programs? Given the self-interest postulate, the natural way to proceed is to look first to those who stand to gain from such programs.

It is obvious that those who will be the direct recipients of transfer programs will see advantages in supporting them politically, and can be expected to do so. It is undoubtedly true that many people may be reluctant, at least initially, to accept charity, either public or private. But most people are remarkably adept at rationalizing the acceptability, indeed virtue, of things that provide them benefits, and transfer payments seem no exception.

It is easy, however, to overestimate the influence of the poor on the size and structure of poverty programs. The political process responds to the concerns of majorities when these concerns are expressed clearly at the polls, and to pressures from relatively small groups that are well organized around a common interest, typically an occupational or professional interest. The poor are in the minority, are less likely to vote than more affluent citizens, and are seldom organized into politically focused groups. It is unlikely, then, that the poor will be very effective in the competition for political influence. Indeed, if the poor had the skills and attributes that are necessary for effective political action, they would not be poor. So while the political demands of the poor will no doubt push in the direction of increasing the funding for poverty programs, the magnitude of this influence will probably be small.

A more effectively organized and politically potent group with a vested interest in expanding government programs to aid the disadvantaged, is made up of those whose employment depends on these programs. These people include the several million people who are employed either directly by government welfare agencies, or who as private contractors and academic researchers assist and advise these agencies. With the job security and income of these people tied to the funding of poverty, and poverty related, programs, they have a common and dominant interest in the continuation and expansion of this funding. When this focused interest is coupled with the relative ease with which people can organize for political action through their employee organizations, poverty professionals become much more effective at obtaining political funding for poverty programs than the poor themselves.

Another group which has a strong private interest in particular types of programs for assisting the poor are those who supply products that it is believed the poor should have but can not be depended upon to buy, even if they had the money. An implicit assumption behind the structure of many poverty programs is that the poor are not responsible enough to make proper consumption decisions. Of the income transfers that are means-tested, provided only if the recipient has a sufficiently low income, only a little more than one-third is in the form of cash payments.[5] The remaining aid takes the form of food, medical attention, housing, educational benefits, and so on. One can be sure that the agricultural lobby, the American Medical Association, along with the construction and education associations, have significant political input into the decisions on the funding of in-kind programs. It should surprise no one, for example, that the incomes of medical doctors are increased by government programs like medicaid which provide poor people with medical care.

The emphasis of self-interest does not imply that people, whether in their public or private capacities, are completely lacking in compassion and concern for others. Indeed, a discussion of the motivations for political action would be woefully deficient if it did not recognize that people do have genuinely charitable feelings for the less fortunate. People not only have these feelings, they also give them political expression at the polls. Furthermore, expressions of charitable feelings will be exaggerated when transmitted through voting behavior. There are two reasons for this. First, our desire to be charitable is like all of our other desires: It is decreased when the cost goes up and increased when the cost goes down. Second, the individual will perceive the cost of making charitable contributions to be lower when made collectively (through voting, for example) than when made privately.

Most people make private donations to charitable organizations because they realize a genuine sense of satisfaction from helping others. But these donations impose a direct cost on the individuals making them, and this cost is taken into consideration when deciding how much to give. As a consequence, when it comes to reaching in one's pocket to support worthy causes, most people are charitable, but not very charitable. Now consider the situation where the decision on how charitable to be is made by casting one's vote. Voting is a very effective way for individuals to satisfy their expressive urge (to make a statement) but an individual vote is unlikely to have any effect on the final outcome. This is simply a matter of arithmetic. In a national election, with tens of millions of votes being cast, the probability that one vote will be decisive is effectively zero. People are aware of this fact and recognize that the taxes they will actually end up paying for public assistance programs is completely independent of how they vote on these programs. The political process removes the conflict between the desire to feel charitable and the personal cost of actually being charitable. The good feelings individuals get from expressing generosity in the voting booth is a real bargain.[6]

---

[5] See Goodman, "Poverty and Welfare," p. 144.

Having discussed the incentives that exist for people to turn to the political process as a vehicle for transferring wealth to the poor, the next question is: How effective at helping the poor can we expect government transfer programs to be?

The Eroding Effectiveness of Government Assistance

Government transfers from the general community to targeted groups such as the poor can obviously make the recipients better off, at least temporarily. The long-run effectiveness of government transfer programs is not so obvious. The reasons for expecting the effectiveness of these programs to decline over time fall under two general headings: first, the dynamics of the political response to transfer programs and second, the recipient and private sector response to these programs.

Political Responses

The political responses to transfer programs, once they are in place, are best understood by considering the previously discussed political motivations for initiating these programs. The important thing to recognize is that the private interests benefiting from the existence of a transfer program will also benefit from, and therefore favor, its expansion. Those who work for the government agencies administering these programs, either directly as employees or indirectly as outside consultants and researchers, will benefit from their expansion. They are more likely to receive promotions, perquisites, job security, prestige, and power if their agencies are expanding than if they are stagnant or in decline.

Transfer agency professionals, then, have a clear incentive to structure their programs in such a way as to broaden their political appeal. This can be done by arranging it so that fairly narrow, but politically organized, interests benefit from an expansion of the program. It is in the interest, then, of government purveyors of benefits to the poor for many of these benefits to take the form of in-kind services provided by politically active occupational groups such as medical doctors, college professors, and farmers. It is also in the interest of the transfer establishment to expand the number of recipients to public generosity as this serves both to increase the size of their agencies and broaden the relevant base of political support.

---

[6] For an elaboration of this point, see Gordon Tullock, "The Charity of the Uncharitable," *Western Economic Journal* 9 (December 1971): 371–92. It should also be pointed out, that just as benevolent feelings can be magnified through the political process, so can malevolent feelings. Just as individuals in a mob are capable of committing atrocities collectively that they would never consider committing as individuals, so individuals may vote for atrocities that few, if any, would favor if their decision were controlling.

The compatibility between the interests of government employees of transfer agencies, the interests of potential providers of in-kind services, and the interests of "deserving," but not necessarily poor, recipient groups is obvious. Once a particular transfer program is in place, those interest groups that the program benefits, or potentially benefits, can be expected to coalesce politically behind it. The effectiveness of these interest group coalitions will, in general, increase with time. Any organizational effort will encounter inevitable start up problems and it will take time and experience to overcome some of the free-rider problems that even quite small groups face.[7] Therefore, the longer a transfer program has been around the more pressure politicians will face to structure and fund it in accordance with the interests of those who are not necessarily poor. Of course, the poor may realize some benefit from this special interest pressure since it will be to the advantage of the special interests to push for, among other things, additional funding for transfer programs. The increased effectiveness over time of the transfer coalition will not be matched by a similar increase in the political effectiveness of the unorganized taxpayers who shoulder the costs of the transfer programs. With these costs widely spread over the entire population they will be almost invisible politically. So once a transfer program has been initiated, the tendency will be for it to expand.[8]

That transfer programs have expanded, there can be no doubt. Medicaid, which provides medical care to the poor, grew to a $27.6 billion program in 1981, up from only $0.5 billion in 1965. Medicare, which provides medical care for the old, grew from nonexistence status in 1965 into a $38.4 billion program in 1981. The Food Stamp program cost $9.7 billion in 1981, a 243–fold increase over its 1965 cost of $40 million. Aid to Families with Dependent Children (AFDC) went from a cost of $1.7 billion in 1965 to a 1981 cost of $12.8 billion, while over the same interval the Housing Assistant program increased from $300 million to $6.6 billion.[9] Not only do transfer programs experience a period of rapid growth, but also, as one would expect with the interest groups that coalesce around transfer programs becoming more influential over time, this growth tends to occur in ways that increasingly benefit these interest groups. For example, with those groups which supply in-kind benefits through poverty programs benefiting by extra demand for their products, organizations such as the American Medical Association and the Mid-American Dairymen will be pushing for an expansion of in-kind transfer programs. Therefore, it should come as no great surprise that while cash transfers to the poor increased 402 percent from 1965 to 1981, in-kind transfers to the poor increased by 5,226 percent over the same period.[10]

---

[7] For a very interesting discussion of this point and its wide ranging implications, see Mancur Olson, *The Rise and Decline of Nations* (New Haven, Conn.: Yale University Press, 1982).

[8] This is not to say that there is no political resistance to the expansion of transfer programs. Obviously there is, or they would quickly expand beyond limit. If a program favored by one interest group is expanded it will, to some extent, do so at the expense of a program favored by another politically influential group. But this conflict can be moderated by pushing back the political budget constraint by imposing more cost on the public. And this is exactly what the politician will be under constant pressure to do.

But are not the poor made better off by expansions in transfer programs, whether the benefits are provided in-kind or otherwise? Very likely the answer is no. The reason for this is that the expansion often results from extending the coverage of a means-tested program exclusively for the poor to one that also provides transfers to the nonpoor. Even if the poor continue to receive the same benefits after such an expansion as before, they are still worse off by virtue of the fact that at least part of the additional benefits will be financed through higher taxes on the poor. Also, once the nonpoor are allowed to compete against the poor for transfers it will commonly be the case that benefits received by the poor are reduced. Those with relatively high incomes have outcompeted the poor in the private sector and there is no reason to doubt that they will do the same in the public sector.

Consider the case of the British National Health Service (NHS) that, in 1948, promised unlimited free medical health care of the highest quality to all. If political promises constituted genuine commitments the NHS should have been no threat to the poor, and indeed should have benefited them. Before 1948 the poor in Britain received medical attention through public and private charitable contributions targeted specifically to the poor. It would be naive, of course, to believe that this medical care was typically of the highest quality. But it also would be naive to accept political promises at face value, and there are reasons for believing that the NHS *reduced* the quality of medical care for the poor. By 1949 the soaring cost of the NHS forced British politicians to face economic reality and impose a ceiling on the NHS budget. Available health care was limited by the budget to less than Britons' wanted at zero price, and therefore had to be rationed by the authorities. Although the political rhetoric suggests that this rationing would be done on the basis of medical need, the reality was, and still is, far different. In obtaining medical attention from the NHS an individual's political influence, community standing, and personal contacts have always been important considerations, possibly more important than the individual's medical problems, and this obviously favors the nonpoor over the poor. This has been recognized even by supporters of the NHS such as the late Richard Titmuss who complained in 1968 that

---

[9] These statistics come from Sheldon Danziger, Robert Haveman, and Robert Plotnick, "How Income Transfer Programs Affect Work, Saving, and the Income Distribution: A Critical Review," *Journal of Economic Literature* 19 (September 1981): 975–1028, Table 1.

[10] These figures were compiled by Goodman, "Poverty and Welfare," p. 145.

> the higher income groups know how to make better use of the NHS. They tend to receive more specialist attention; occupy more of the beds in better-equipped and staffed hospitals; receive more elective surgery; and better maternity care, and are more likely to get psychiatric help than low income groups—particularly the unskilled.[11]

The evidence suggesting that the poor in Britain have been harmed by the expansion of transfer payments that took place through the NHS is rather dramatic. The death rate of the poorest segment of the British population actually increased after the NHS went into effect, and this was at the time medicine was being revolutionized with the introduction of antibiotics and the overall death rate was dropping.[12] Furthermore, the death rate for the least advantaged in Britain continued upward. The male mortality rate for unskilled workers was 16 percent higher for the years 1970–72 than it was for the years 1949–53. This contrasts with decreasing mortality rates for all other classes of British males over the same period. For example, over the interval 1949–53 unskilled male workers had a mortality rate 37 percent higher than did professional male workers, but by 1970–72 the mortality rate of unskilled male workers was 78 percent higher than that of professional male workers.[13]

---

[11] As quoted in *The Economist,* 28 April 1984, p. 19.

[12] See Gordon Tullock, *Economics of Income Redistribution* (Boston: Kluwer-Nijhoff Publishing, 1983), pp. 100–1.

In the United States government transfer payments for medical care have expanded rapidly in recent years as the program targeted to the poor, Medicaid, has been joined by Medicare, which is a non-means-tested program for those over 65. This expansion has almost surely made the poor worse off as they now have to pay taxes to provide medical care for the nonpoor. Those over 65 are, on average, wealthier than those less than 65, and the Medicare program is larger and has grown more rapidly than the Medicaid program.[14] Also, because the Medicare and Medicaid programs have become much more expensive than anticipated when first established, measures to eliminate their expenditures have been seriously considered. If such limits are imposed the political competition between the poor and the nonpoor will become much more direct and those who are most disadvantaged will surely become more disadvantaged.

Another program that was originally aimed at assisting the less affluent is the student loan program which subsidizes students by providing them with low interest loans. Until 1978 these loans were seldom available to a student whose family earned more than $25,000 a year. In 1978, however, Congress yielded to the influence of the higher education community and expanded the student loan program to include those from high income families. By 1982 the volume of student loans had increased by a factor of five.[15] The subsidy component of this program was further increased by the fact that the federal government has been lax in making sure these loans are repaid. At one point in 1982 the default rate on student loans was 12.5 percent. A report by the General Accounting Office showed that 6,000 doctors had defaulted on student loans totaling $5.2 million. One doctor who had reportedly received $623,000 in Medicaid payments had not repaid a federal loan for medical school of $4,750.[16] In addition to loan subsidies the federal government has also transferred direct financial aid to students, and those who are by no means poor receive a major portion of this aid which totaled over $6.2 billion in 1981. After the taxes of the poor are taken into consideration, it is by no means clear that as a group they receive positive net benefits from these student aid programs.

---

[13] These figures come from *The Economist*, 28 April 1984, p. 19, Table 2.

[14] For a detailed discussion of how Medicare discriminates against the poor, see John C. Goodman, "Solving the Medicare Crisis," Cato Institute *Policy Report* (February 1984): 6–9.

[15] Jane Bryant Quinn, "The Student Loan Scare," *Newsweek*, 24 May 1982, p. 68.

[16] Dennis P. Doyle, "The Federal Student Aid Mess," *Wall Street Journal*, 19 May 1982, p. 28.

Private Responses

Over time private responses to the existence of government transfer programs tend to weaken the effectiveness of these programs at reducing poverty. It seems reasonable to expect that as government increases the amount of money it transfers away from the general taxpayer, in order to supposedly help the poor, the amount that people contribute to private charity will decrease. And indeed there is evidence to support this presumption. Using 1948–72 data economists Burt Abrams and Mark Schmitz found that for every dollar increase in government welfare expenditures, private charity was reduced by 28 cents.[17] But other studies have suggested that the private charitable contributions have actually increased slightly as public transfers have gone up.[18] One could explain this by arguing that higher government transfers have resulted in higher marginal tax rates and this, of course, lowers the cost of tax deductible charitable contributions. So it appears that the affect of government transfers on private charity is ambiguous.

It is important to recognize, however, that the above studies follow the Internal Revenue Service in their definition of private charity. And, U.S. tax law allows contributions that have little if anything to do with helping the poor to qualify as tax deductible charitable contributions. If the composition of these charitable contributions is examined one finds that as public transfers began increasing in the 1930s private giving shifted significantly away from helping the poor and into the support of religious organizations, education, health, and the arts. In his study of the effect of public transfer programs on private giving, economist Russell Roberts concluded, "The huge growth in public transfers in the 1930s crowded out private antipoverty efforts and fundamentally changed the nature of private charity."[19]

Government transfer programs also impact on the decisions made by recipients, and when these recipients are poor they are often motivated to make decisions that will, over the long-run, keep them poor. The development of human capital and employment skills is the best way to rise above poverty. Any program that discourages those who are poor from actively seeking out work discourages the development of productive skills and serves to perpetuate poverty and dependency. This fact is at the heart of one of the major problems with poverty programs since these programs, in effect, impose very high taxes on earned income.[20] While most people associate high marginal tax burdens with high incomes, it is in fact the poorest in our country who pay the highest marginal tax rates. The recipients of AFDC, for example, lose a dollar in benefits for every dollar earned (and known to be earned by the authorities). This marginal tax rate of 100 percent can actually understate the marginal tax burden facing the poor because of the possible loss of some in-kind benefits, such as Medicaid, that results as earned income increases beyond certain thresholds.[21] It is not hard to predict that a response to marginal tax rates of this magnitude will be less than an ambitious effort to acquire work experience and job skills. So once an individual becomes a recipient of a program like AFDC the temptation is to respond in ways that reduce future income earning potential and increase dependency on the program.

---

[17] Burt Abrams and Mark Schmitz, "The 'Crowding-Out' Effect of Government Transfers on Private Charitable Contributions," *Public Choice* 33 (1978): 28–40

[18] Robert Schwartz, "Private Philanthropic Contributions: An Economic Analysis," Ph.D. dissertation, Columbia University, 1966.

[19] Russell D. Roberts, "A Positive Model of Private Charity and Public Transfers," *Journal of Political Economy* 92 (February 1984): 147.

Not only does a program like AFDC motivate a tendency toward dependency on the part of the recipient, but there is a more insidious tendency for this dependency to be passed on from one generation to the next. For many disadvantaged teenage girls the opportunity to benefit from AFDC and other social programs is the most attractive financial option they have. All they need do to take advantage of it is to have a child out of wedlock. Even if AFDC payments do not motivate an intentional decision to become pregnant, the availability of these payments clearly reduces the incentive to be careful. While marriage is always a possibility it is often not a financially attractive substitute for AFDC payments. Most mothers of illegitimate children do eventually marry but typically these marriages do not last long. As George Gilder has vividly pointed out, a low income mother will find indexed welfare benefits "more preferable in everyway to the taxable, inflatable, losable, drinkable, druggable, and interruptible earnings of a man."[22] The degree of substitutability between men and transfer payments can be debated, but the fact remains that as AFDC and other transfer benefits have grown so has the percentage of illegitimate babies born to poor mothers. It is true, of course, that many things influence the rate of illegitimate births other than the availability of transfer payments for unwed mothers. But in one study, based on a sample of unwed pregnant teenagers, it was concluded after statistically controlling for such things as age, religion, school enrollment, and ethnic origin, that those eligible for public assistance were significantly more likely to deliver their baby out of wedlock.[23] When this is coupled with the fact that children born out of wedlock appear to be neglected and abused with higher probability than normal,[24] then it is reasonable to surmise that welfare dependency in one generation tends toward welfare dependency in the next.

---

[20] For evidence that suggests strongly that the labor-force participation rate for the poor has been reduced by the growth in social welfare programs, see Charles A. Murray, *Losing Ground: American Social Policy 1950–1980* (New York: Basic Books, 1984), chap. 5.

[21] See U.S. General Accounting Office, "An Evaluation of the 1981 AFDC Changes: Initial Analyses," PEMD-84-6, 2 April 1984, p. 7.

[22] George Gilder, *Wealth and Poverty* (New York: Basic Books, 1981), p. 123. For evidence on the apparent connection between the increased breakup of poor families and social welfare programs, see Murray, *Losing Ground.*

There is direct evidence to support the depressing view that families are being acculturated into persistent welfare status. According to one study, the group most likely to become AFDC recipients are those whose mothers were recipients.[25] A 1967 study by the Department of Health, Education and Welfare concluded that just over 20 percent of those AFDC mothers whose welfare history was known had received welfare as children. There is no mention of how much this figure has grown, if any, in subsequent HEW studies, but it has been estimated that "trans-generational recipients" constituted up to 40 percent of all AFDC mothers by 1980.[26]

It is not at all clear that the income welfare recipients receive from public transfers is any more than many of them would have been earning had they been required to take responsibility for the wealth creating consequences of their decisions. Indeed, it is surely the case that for many individuals trapped into welfare dependency they would have been better off financially had they faced stronger incentives to develop income earning skills. Through either conscious decisions or the subtle influence of acculturation, untold numbers have simply substituted publicly provided income for privately earned income with the net effect of public transfers on their incomes being close to zero or worse.

---

[23] See Victor R. Fuchs, *How We Live: An Economic Perspective on Americans from Birth to Death* (Cambridge: Harvard University Press, 1983), pp. 105–6.

[24] Ibid., p. 28.

[25] Morley D. Glicken, "Transgenerational Welfare Dependency," *Journal of Contemporary Studies* 4 (Summer 1981): 31–41.

[26] Ibid., pp. 28–29.

The Distribution of Income

We have identified and discussed some specific reasons why the ability of government transfer programs to help the poor should be expected to weaken over time. Although supporting evidence from particular programs has been presented, the important question is whether or not the overall distribution of income has become more equal in response to the enormous amount of money that has been spent on poverty programs. Based on the evidence that exists, it appears that the impact of the huge growth in transfer programs has been rather minor. In a study of the distribution of income and wealth in the United States Reynolds and Smolensky concluded that, when the net benefits of all government programs were added to the incomes U.S. households received from labor and capital, there was no significant change in the distribution of income between 1950 and 1970. For example, according to their study, those households in the lowest 20 percent of the income distribution received 6.4 percent of national net income in 1950 and 6.7 percent in 1970. At the other end of the distribution, the percentage of net income going to those households in the top 20 percent declined slightly from 39.9 in 1950 to 39.1 in 1970.[27]

Given the increase in transfer expenditures that occurred over this period of time this is a rather startling result, but one quite consistent with our analysis of the political incentives that create and are created by government transfer programs. In providing their own explanation for the impotency of these transfer programs Reynolds and Smolensky state:

> The very scale of government spending programs today mitigate against augmenting their capacity to redistribute net income. . . . Even if a democratic government initially directed additional benefits and costs "efficiently" toward an identifiable group, it would find efficiency hard to preserve. With such large sums at stake, more players find it worthwhile to compete for a larger share of the benefits and a smaller share of the taxes.[28]

More recent studies of the distribution of income have reached conclusions similar to those of Reynolds and Smolensky. For example, evidence cited in an extensive review of the literature on the affect of U.S. income transfer programs, indicated that over the period 1965 to about 1978 the "income inequality has remained relatively constant."[29]

---

[27] Morgan Reynolds and Eugene Smolensky, "Distribution of Income and Wealth: The Fading Effect of Government on Inequality," *Challenge* (July/August 1978): 32–37.

[28] Ibid., p. 36.

[29] See Danziger et al., p. 978.

It should be pointed out that some studies indicate that government transfer programs have resulted in an income distribution that is significantly more equal than it would otherwise be.[30] The general approach of these studies involves comparing the post-transfer distribution of income with the counterfactual income that would have existed in the absence of government transfers. Unfortunately, estimating "what would have been" is extremely difficult and typically the comparison is between the income distribution with government transfers and the income distribution that results when these transfers are eliminated. But obviously this overstates the effect of the transfers. As previously discussed, the response to the availability of public transfers has been a reduction in private charity and the substitution, on the part of the poor, of transfer income for earned income. To simply eliminate the effect of transfer programs after people have, over a number of years, adjusted their behavior to them yields income estimates for the poor that are lower than the income they would have received, either from private charity or earnings, had the public transfers never been available. Studies, such as the one by Reynolds and Smolensky, which look at the income distribution, as it actually is, over time and find that it has not changed appreciably in response to large increases in government transfer programs, speak more accurately to the question of how effective these programs have been than do studies which see the income of the poor increasing a dollar for every transfer dollar received.

In discussing the income distribution we have been talking about the effect of government transfer programs on the *relative* income of the poor, as opposed to the absolute income. If, as the evidence seems to indicate, massive transfer expenditures have not increased the relative income of the poor, then it is surely the case that these expenditures have decreased the absolute income of the poor. The taxes necessary to finance transfer payments reduce the private returns to labor, to savings, and to investment, and to wealth producing activity in general. Therefore these taxes create wealth-reducing distortions in the economy. Similarly, the payment of transfers typically reduces the incentive of recipients to engage in productive activities. A large number of empirical studies have been made in the attempt to estimate the size of the negative productivity effects of taxes and transfers. Not surprisingly these estimates vary, with some studies finding large negative effects and others finding much smaller effects.[31] But no one has seriously argued that the effect of taxes and transfers on economic productivity is neutral or positive. And even if the huge increase in transfer programs over recent decades, and the huge increase in taxes necessary to finance them, have had only a slightly negative effect on national income, these programs have still made the poor absolutely worse off if they have left the distribution of national income unchanged. If the negative impact of taxes and transfers on productivity is as large as many believe to be the case, then the poor could be left absolutely worse off even if their relative income has been increased by government transfer programs.

---

[30] These studies are discussed in Danziger et al., pp. 1006–15.

[31] Many of these studies are discussed in Danziger et al.

None of this is to deny that transfer programs can, and indeed do, benefit particular groups, including the poor, in the short-run. But these short-run benefits come at the expense of long-run reductions in the general economic welfare that harm the poor as well as the nonpoor. But it is the short-run consequences of transfer programs that dominate myopic political considerations, not the long-run consequences. Furthermore, the eroding effectiveness of government transfer programs not only fails to discourage politicians from embracing these programs but has the perverse effect of actually increasing the political support they receive.

The Political Success of Economic Failure

In order to understand the political success of economic failure in the case of transfer programs, it is useful to consider the argument most commonly used by those who believe these programs have been successful in helping the poor. This argument relies on statistics which indicate that if the anti-poverty programs were eliminated the percentage of the population falling below the official poverty line would increase dramatically. For example, figures for 1982, published by the Congressional Research Service (CRS), indicate that in the absence of government support programs the poverty rate would have gone from 8.8 percent to 24 percent. Also according to calculations by the CRS, the elimination of government programs would result in 14 to 16 million elderly people, between 55 and 60 percent of the elderly population, living in poverty.[32]

Little thought is required to recognize this argument forms a dubious basis for heralding the success of government transfer programs. If these programs had never come into existence it is inconceivable that one-fourth of the entire population, or over one-half of the elderly, would be living in poverty. With the data indicating that the increase in social welfare programs have had little impact on the income distribution, the rate of relative poverty in the counterfactual world with no history of government transfers would be little different than that which is actually observed with these transfers. The argument that the elimination of transfer programs would result in high rates of poverty points to the failure of these programs, not their success. The programs are doing little to reduce the poverty rate below what it would be in the long run without them, but they are doing a lot to discourage independence and encourage dependence. This is hardly a testimony to the success of these programs. It is, however, the reason these transfer programs are so resilient politically.

---

[32] These figures are cited by Spencer Rich in his article, "The Skeptics Are Wrong: Anti-Poverty Programs Do Work," *Washington Post*, 6 May 1984, p. Fl.

Even if transfer recipients would have been better off in the absence of transfer programs, having accommodated to these programs they would experience genuine suffering in the short run if the programs were suddenly scaled back or eliminated. Welfare recipients cannot develop instantly the skills and earning potential needed to replace welfare income. From the perspective of individuals who have failed to develop productive skills, who have had children they cannot themselves support, or who have absorbed an attitude of dependency from their surroundings, the short run may extend over an extensive period of time. Certainly, we are talking about a period of time that extends well beyond the time horizon of the political process. The political pressures that would be brought to bear on any serious attempt to curtail transfer programs, whether they are targeted to the poor or otherwise, would obscure the long-run benefits that would result from such a curtailment. Once the political process starts down the transfer path it is no easy task to turn back, even after it has been determined to be the wrong path.

Instead of reversing direction and reducing transfer payments, the political process will more likely be pushing in the opposite direction. The failure of transfer programs will not be seen to be inherent in the politics of these programs. Instead, continuing poverty will be seen as justification for an expansion of existing transfer programs and/or the establishment of additional programs. This view is reinforced by political myopia. Measured poverty can be temporarily reduced by another infusion of government transfer payments, and near-term political gains are available to politicians who are able to take credit for such a reduction in poverty. Of course, over the long run the expansion in transfer programs will fail to reduce relative poverty, and will likely increase absolute poverty. But with the future consequences of political action being discounted as heavily as they are, these future failures will be largely ignored.

The further the political process travels down the path of ever larger transfer programs, the more difficult it will be to reverse course.

As more people respond to the expansion of transfer programs by becoming dependent on them, the short-run political cost of scaling these programs back becomes greater. The greater the failure of social welfare programs, as measured by the dependence they create, the more successful they will be, as measured by political viability.

Conclusion

Not only are an increasing number of poor people in the country caught in a trap of welfare dependency, we as a society are also snared in a welfare trap. Having created a dependent subculture we cannot, indeed should not, suddenly cut these people off. To do so would be to practice an extreme form of political deception and cruelty.

But we can at least be realistic in assessing the effectiveness of government welfare programs. It is surely not very useful to continue with the pretense that these programs have been successful when, in fact, they have been failures. At the very minimum, we should recognize the folly of attempting to overcome the failures of existing programs by further expansions in these programs.

Yet if we continue to trust in politics as usual, it is almost certain that transfer programs will continue to grow and proliferate. Electing the "right" politicians, urging "sensible" reforms, or swings in the public attitude may temporarily slow this growth on occasion, but ever present political incentives will always be pushing toward expansion. And the only natural barrier to this expansion is economic stagnation and decay.

The best, possibly the only, hope for halting the destructive excesses of ever increasing transfer programs is through a collective realization that we all suffer, rich and poor alike, when the government is free to hand out benefits to every group which is prepared to organize politically around some narrow interest. Overriding the myopic pressures of ordinary politics with some type of constitutional cap on the percentage of our national income that government can spend would be a beneficial result of this realization. That constitutional amendments of this type are being discussed and considered seriously today is an encouraging sign.

Of course, many will argue that any arbitrary constraint on government will hurt the poor. But this is simply not true. To recommend limiting government expenditures is not to lack compassion and concern for the poor. Quite the contrary. Once one faces up to how the political process actually works, as opposed to how we might ideally wish it worked, it is realized that both compassion and common sense will be served by constraining the size and scope of government.

THE POVERTY OF NATIONS

Karl Brunner

The Condition of Poverty and the Creation of Wealth

Poverty remains an endemic state of man. The curse imposed by an angry god on Adam and his descendants describes the human fate. Man experienced over history, with rare exceptions, toil, hardship and oppression. The uncertain and stringent conditions of life challenged man's awareness. He sought for answers explaining this fate and these were couched in the form of myths and legends. An imaginative human mind invoked a fall from grace or a curse imposed by gods. Such orientations naturally affected man's expectations of liberation from the grinding burden of toil and hardship. The restoration of grace appeared to be the crucial condition for such liberation.

This elementary message emerged over the centuries in very different forms and in widely varying detail. The basic story of man's fate conveyed by ancient myths still persists with a subtle and pervasive influence. The dominant ideologies of our time promise the restoration of (at least secular) grace with the liberation from human bondage to poverty. This promise of liberation should be realized by a collective action imposing a set of explicitly designed *political* institutions as the dominant mechanism for the coordination of society.

The ancient themes still reverberate under new labels in various branches of sociology or social psychology. But their influential repercussions on the intelligentsia market have been challenged for 200 years by the evolution of economic analysis. This analysis, initiated by the Scottish philosophers of the 18th century, offered a revolutionary insight into the social context of man's life. This insight

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Fred H. Gowen Professor of Economics and Director of the Center for Research in Government Policy and Business at the Graduate School of Management, University of Rochester. An earlier version of this paper appeared in the January 1985 issue of *Business Economics*.

has hardly been absorbed or understood by the educated middle classes of western societies and even less by the professional articulators.

The Scottish tradition of economic analysis explains the emergence of social groups and "great societies" extending social interaction far beyond face-to-face associations. It also explains the emergence of political structures without invoking metaphysical entities. This tradition shows in particular how the social structure emerges as resourceful human agents interact to improve their lot according to their judgment and understanding. The social structure appears in this manner as an important but quite unintentional by-product of this interaction.

The same analysis also addresses the condition of poverty and the creation of wealth. The wealth of nations, expressed by the standard of living, depends foremost on individual effort, ingenuity, and imaginative adaptation to the environment. The natural environment together with the resources provided by nature can make a substantial difference for a nation's opportunities. These opportunities offered by nature condition the pattern of activities and the use of available human and nonhuman resources. But the possessions of natural resources do not determine a nation's wealth and do not, per se, suspend Adam's curse. Nature's heritage forms, contrary to widely held beliefs well represented by the Brandt Report on the North-South dialogue, neither a necessary nor a sufficient condition for high or rising levels of wealth. More than 50 years ago Argentina was expected to evolve on the basis of its natural resources into one of the world's wealthiest nations. It approached in contrast a pattern of permanent stagnation. We still encounter "experts" informing us that Argentina with its vast resources should really be expected to overcome, in due course, its current difficulties. Brazil should also be on this count among the wealthiest nations today. Switzerland, less favored with natural resources than Zaire, Nigeria, Mexico, and other nations, should appear among the poorer nations appealing for aid from Argentina, Brazil, Mexico and others. We may also compare Hong Kong, South Korea, Taiwan, and Singapore with more favored nations. Or consider the permanent stagnation of Tanzania and Zimbabwe in Africa or of Bolivia and Peru in South America.

Political power has frequently been singled out as a crucial factor of a more widely conceived "natural environment" which decisively affects the wealth of nations. The Brandt Report exemplifies once again this view. But history and economic analysis fail to support this claim. Nations with little power experienced over the past 100 years or more a remarkable increase in wealth. This accrual of wealth occurred moreover not at the cost of other (more powerful?) nations but contributed actually to benefit others through an expanded economic interaction.

Poverty is characterized by a low rate of production of goods and services. The creation of wealth involves an expansion of this production. Western nations experienced this process over the last 200 years on a scale never before recorded in history. Power can simply not explain this phenomenon. Power does not produce goods, it simply absorbs and uses goods to maintain the power apparatus—a *nomenklatura*. Power, however, may be applied to extract wealth from those who produced it, an exercise that does not create wealth but actually destroys it in the final analysis. The emphasis on power thus confuses the *creation* and the *redistribution* of wealth. It also fails to understand the longer-run consequences of redistributive power. The argument based on power frequently occurs in a context linking the accumulation of wealth in Western nations with colonialism. But Adam Smith already elaborated the crucial implications of colonialism. He recognized that colonialism meant a loss in wealth for the colonizing nation combined with a domestic redistribution favoring specific groups involved in the process. This redistribution, and not any general accrual of wealth, was the historical driving force of colonization. The exercise of power applied to colonial efforts thus imposed an economic burden on the mother country.

Our attention clearly must be directed beyond natural resources and power to human effort and resourceful ingenuity. The founding fathers of modern economic analysis recognized the central importance of human capital and investment in human capital as a condition of the wealth of nations. The development of the German economy after World War II dramatically confirmed this contention. But effort and resourceful ingenuity are not "sociological data" determined by the mysteries of a "mental attitude," a "religious commitment" or an inherited "work ethic." Customs, traditions, and cultural values may play a role for a while. However, they will diminish in importance and behavior will adjust once the supporting conditions have eroded. Economic analysis informs us that effort and resourceful probing emerge with substantially greater frequency and intensity when human agents expect to capture the benefits of their search activity and their ingenuity. Without the incentives of potential rewards agents hardly will find it worthwhile to expend much effort and ingenuity. The magnitude and quality of effort and the intensity of search for imaginative innovations in types and use of resources is generally closely linked with the expectation of capturing the returns from these endeavors. The penalty-reward system thus conditions the current level of human endeavor. Economic incentives also shape the accumulation of human capital expressed by the level and quality of skills and the development of nonhuman resources with the productive modification of the environment.

The power argument outlined above fails to comprehend the central role of incentives and the close link between the investment of one's efforts and the capture of expected returns. It proceeds as if there operated no incentive feedback from the distribution of the product to its supply. The argument suggests on the contrary that all goods are to be grabbed by participants in a social game. The powerful acquire the lion's share of the economic pie and the losers get the crumbs. The natural conclusion from this romantic vision is that the balance of power needs to be changed.

The message conveyed by economic analysis thus directs our attention to the conditions that encourage the application of human effort and ingenuity to the inheritance determined by history and the environment. These conditions do not depend on nature. They are determined by the social organization characterizing a society. The prevailing sociopolitical institutions condition the potential opportunities facing human agents. They ultimately shape the incentives that guide human effort and ingenuity, and that evolve over time into customs, habits, and traditions. Such customs and traditions, however, can be destroyed by institutional changes that lower the expectation of a link between effort and real benefit. Thus we recognize in the pattern of existing sociopolitical institutions the crucial condition for both wealth creation and the persistence of widespread poverty. A stagnant economy and persistent poverty therefore do not express a curse of vengeful gods (in the heavens or in history) to be atoned by a purifying collective action. Rather, such stagnation and poverty result essentially from the inheritance of social institutions affecting individual behavior.

The Ambivalence of the Political Structure

The nature of institutions maintaining poverty or fostering wealth creation requires some further attention. The role of the political structure, and specifically of the state, needs to be clarified in this context.

The social productivity of the political structure is best understood in comparison with a state of anarchy. An agent arranging his affairs under anarchy has the following options for the use of his resources: he can invest in production, in trade, in attempts of robbery, and in defense against others. The exposure to potential loss of resources due to piracy by others lowers the incentive to invest in production. It also constrains the opportunity to trade. Under the circumstances productive activity addressed to wealth creation remains at a low level. All agents seem to be caught in the trap of a prisoner's dilemma. The repetitive occurrence of the basic social problem posed by anarchy, however, induces the interacting responses so well described by the founders of economic analysis and more recently elaborated by Hayek (1982). The evolution of political institutions offers in this respect a solution to the prisoner's dilemma inherent in a state of anarchy.

Consider the options available to an agent arranging his activities in the context of the political structure. He can invest his resources again in four different ways: in production, in trade, in the political process to redistribute wealth on his behalf, and in the political process as a defense against wealth redistribution schemes advanced by others. Anarchy and the political structure are thus not distinguished by presence or absence of a zero (or possibly even negative) sum game of social interaction. We recognize at this stage the peculiar ambivalence of the political structure. This structure establishes rules of the social game. Such rules substantially confine the wealth-impeding activities of anarchy but they never totally eliminate them, as exemplified by the Mafia and the Camorra. The rules therefore provide a "monopoly of violence" anchored by the political structure. This monopoly forms a necessary condition for the specification of property rights and their enforcement. Investment in production and in activities raising the level and quality of human and nonhuman resources is encouraged as a result of the greater expectation of capturing the returns under the rules of the social game. Specification and enforcement of property rights also encourage a wider range of possible transactions and provides new opportunities for mutually beneficial trade. The resulting increase in the creation of wealth expresses the remarkable productivity of a stable political structure that allows individuals to capture the benefits of increased economic efficiency.

The magnitude and extent of the benefits accruing from the political structure, however, depend on the institutional arrangements guiding the coordination of socioeconomic activities. In this context private property exhibits a crucial advantage that is not well understood by most professional articulators in the public arena. The assignment of property rights resolves a physical impossibility associated with potential social conflicts. Scarce resources cannot be controlled simultaneously by several persons. The structure of rights determines who can do what with respect to which scarce object.

This structure fully reflects the inherited patterns of resources. It mirrors in this sense the constraints imposed by nature. All resources need be owned by some person and beyond this no rights should be assigned.[1] This condition rules out the wide array of market closing and entry restricting arrangements under any name. Private property rights also guarantee the important link between effort and reward that was emphasized above. Under such a system this link is actually generalized to transactors beyond the range of owners of nonhuman resources.

No rules of the social game however can preclude a new form of wealth-impeding activities from replacing the ancient patterns of anarchic wealth impediments. The problem actually adheres to any set of rules. All political structures determine potential opportunities for manipulation *within* the rules accepted for their operations. They unavoidably offer incentives to be used for purposes of wealth redistribution among the participants of the social game. There is hardly a political institution that does not have consequences for the distribution of wealth. Agents respond to this fact by investing resources in the political process in order to generate wealth transfers from others or to ward off attempted transfers by others. This aspect of the political structure involves basically a negative sum game within the context of the socially productive positive sum game provided by a stable set of rules. The incentives to invest in the political process for purposes of acquiring wealth from others or for protective political actions lowers the allocation of resources to socially productive activities. The taxes imposed by implicit or explicit wealth transfers occurring in one form or another lower moreover the incentive to invest in production, trade, and the accumulation of resources. These consequences determine the inherent ambiguity of the state expressed by the joint operation of a positive and negative sum game proceeding in the context of political institutions organizing the social interaction.

All political structure thus involves simultaneously a wealth-creating and a wealth-impeding dimension. In contrast to the state of anarchy, however, the wealth-obstructing activities proceed in accordance to a recognized and generally accepted set of rules. Political structure thus lowers, but does not remove, the uncertainty confronting agents' socially productive activities. The magnitude of the lowered uncertainty or, in other words, the extent of the wealth-impeding range of activities associated with the political structure depends crucially on the detail of the sociopolitical institution. Every set of political institutions produces its specific mix of positive and negative sum social games. The weight of wealth-impeding activities depends thus on the sociopolitical arrangements of nations.

---

[1] See Meckling and Jensen (1980).

The nature of wealth-fostering and wealth-impeding political institutions may be usefully described in rough outline. The distinction between the two dimensions of the state may be usefully introduced for this purpose. These dimensions refer to the "protective state" and the "redistributive state." This classification is essentially justified in terms of the purpose of political institutions but holds only approximately in terms of their consequences.

Robert Nozick (1974) and James Buchanan (1977) demonstrated the redistributive dimension unavoidably embedded in the protective state. The protective state encompasses a set of political institutions that define general rules offering a stable framework for agents' productive activities. Such rules involve most particularly the definition of private property, the protective arrangements of police and courts associated with this definition, enforceability of privately negotiated contracts, and stable and predictable fiscal procedures including monetary policy. The rules bearing on the government's financial affairs need explicit recognition in this context. Unpredictable explicit changes in taxation or implicit changes via erratic inflation in a world of tax rates addressed to nominal values can substantially lower the incentives of the positive sum social game. A rich variety of observations informs us that the detail of the political framework shaping the social institutions listed above substantially influences the productivity of the social game. Well-designed political structures foster the evolution of markets and improve their functioning. Moreover the higher level of predictability tends to encourage the accumulation of resources. These consequences are essential strands of a wealth-creating social process. This follows from the fundamental fact that the social institutions protected by the type of political structure under consideration assure, on average, a systematic link between effort and ingenuity applied on the one side and the resulting returns on the other. We may formulate this more generally as a set of institutions that raise the probability of a link between the consequences of actions and the agents committing these actions. They also lower the likelihood of market closing and entry restricting activities.

An alternative set of rules or institutions embodied in political structures represented by the redistributive state typically produce wealth-impeding consequences. The class of these political institutions exhibits a rich detail testifying to man's inventiveness. We may divide this vast detail into two broad groups for our purposes. One group involves constraints on choices bearing on contractual arrangements and the tenure of property. The other group contains structures with direct distributional consequences and associated distortions of incentives. Although this division is neither neat nor clean, a better organization of the material may emerge from further discussion.

The argument outlined above suggests that the creation of wealth or entrenched poverty is substantially conditioned by both levels of the political apparatus, the protective state and the redistributional state. Too little of the first and too much of the second obstruct wealth creation and maintain pockets of poverty or even mass poverty. An uncertain or more or less deliberate failure to exercise the basic protective function erodes the link between *productive* effort and capture of expected returns. This experience can be observed over history and continues in our time. I have already mentioned Argentina, and we can point to the permanent political instability and longer-run uncertainty bearing on important sociopolitical institutions in Bolivia, Peru, possibly Chile in South America, Zimbabwe and other nations in Africa. The social context of these nations provides few, or weak, incentives to create wealth through productive investment. The context spurs incentives directed toward the conquest of political control with the acquisition of the associated spoils. The resulting pattern of entrenched poverty cannot be alleviated by doses of foreign aid. Such aid essentially contributes to raise the resources available to the local *nomenklatura*. It cannot replace or offset the absence of an adequate protective state.[2]

The realization of the basic protective function may be offset to some extent by the state's redistributional activities. Constraints on contractual arrangements and the effect of sociopolitical conditions on property tenure exemplify this aspect. The usual textbook treatment of production obscures this important point. We read that output is linked with an array of inputs via a production function. This function purportedly represents the underlying technology. This interpretation is thoroughly misleading, however. The nature of the production function is sensitively conditioned by sociopolitical circumstances. The production function actually forms the outcome of agents' optimizing responses to these conditions. It follows that variations in the admissible range of contractual arrangements or associated organizational forms modify the production function. Constraints on admissible contractual arrangements tend to lower output for any given input. *Consequently they impoverish a society and* obstruct the imaginative search for new modes of wealth creation.

---

[2] See Bauer (1978).

The imposition of self-management or codetermination illustrates the issue. The organizational forms may survive successfully in competition with the corporation and other forms. They may exhibit a comparative advantage for specific activities. Our experience clearly indicates however that this comparative advantage occurs over a very limited range. A coercive imposition of these forms forces, therefore, the organization of production into a mold less adapted to the coordination of resources in production. A comparative decline in output thus emerges. Detailed investigations of the incentives prevailing under self-management and codetermination confirm the results of economic analysis. These investigations also show that self-management in particular creates new social tensions and conflicts between younger and older workers, or between established workers and potential new workers entering the market.

Corporate governance is another fashionable issue that should be mentioned here. Modifications of governance are proposed in order to achieve purportedly desirable social goals. The sense of these proposals usually conveys that they impose *no* social costs and *will* achieve the desired effects. Economic analysis disillusions us on both counts. The social costs in particular will rise with the severity of constraints controlling corporate organizations.

The change in the law governing insolvency in Germany and the "comparable worth" movement in the United States offer further examples of contractual constraints. The assignment of priority rights to several months salary beyond the date of filing for bankruptcy raises the cost of capital, impedes investments, and affects employment. A systematic application of "comparable worth" procedures would seriously impede the organization of production and impose social costs in terms of lessened output of goods and services.

Pharmaceutical regulations and the admissible range for the forms of property tenure are the last examples of constraints on the production function potentially lowering welfare. The prevailing regulation of pharmaceutical products confines both contractual procedures and the choice of production processes. It has increased the cost of operation and, via the reduction in the relevant economic duration of the patent, it has lowered the expected return. Moreover, the regulatory constraints have lowered new pharmaceutical innovation and thus the achievable state of health.

The modification of the patent's relevant duration actually abrogates established property rights. Constraints on the form of property tenure can significantly impair the nature of the social production function. In many countries around the world land is frequently tended by peasants under a usufruct system. Irrespective of the political motives behind this tenure system an examination of its operation reveals that it obstructs productivity of agricultural labor, obstructs investment that would improve the quality of land, and encourages population growth in the countryside. The collectivization of agriculture offers a particularly emphatic example of obstructive institutionalization. It eroded incentives to produce and raised incentives for a wasteful use of resources. Agricultural output naturally suffers. Socialist nations from Tanzania to Russia exemplify our case with remarkable clearness. The land tenure system imposed assured a permanent agricultural crisis.

The second strand of the redistributional state encompasses examples from the welfare system of Western societies, the oligarchic power structure of nations in the Third World, the *nomenklatura* of socialist countries, or some patterns of nonsocialist dictatorships. The welfare system of Western nations imposed a massively accelerated redistribution which lowered incentives to work, invest, and save. On the other hand it raised incentives to invest resources in the negative sum game of political processes.

The oligarchies of many nations in the Third World depend for their survival on a persistent redistribution of wealth from the countryside to the cities. Their political base is usually anchored with the populace in the cities. In order to maintain their power the ruling oligarchies find it advisable to impose low prices on agricultural products. This pattern destroys incentives to produce food, erodes incentives to invest resources in agricultural operations, and creates incentives to abandon fertile land and join the masses in the cities.

Most socialist countries operate a vast system redistributing wealth from the potentially productive sector to the *nomenklatura*—a socially unproductive sector composed of a huge military complex, an internal security and economic control apparatus. Once again this redistribution system lowers incentives for the productive use of resources. Nonsocialist military dictatorships also require for their survival a redistribution favoring the military apparatus with similar consequences for the wealth-creating dimension of the social process.

The evolution of labor markets over the past 16 years reveals that Western nations participated in their own way to obstruct, usually with the best intentions, the wealth-creating process. Unemployment in European countries stayed very low until about the end of the sixties. The pattern changed dramatically during the seventies. Unemployment rates rose to the double-digit range. Belgium and the Netherlands measured recently unemployment rates above 15 percent. With the exception of Switzerland almost all European nations experienced massive increases in measured unemployment. Politicians and the media frequently attribute this development to new technologies. In the public arena one also encounters comparisons with the Great Depression. This comparison fails however. The rising trend in European unemployment occurred, in contrast to the thirties, during a period of expansionary financial policies. Moreover, as we learn from economic analysis, technology cannot explain this unemployment either. Technological innovations change opportunities. Old jobs disappear and new jobs emerge. The American experience of the last 30 years demonstrates that reasonably functioning market processes continuously create new jobs and offer new employment.

The crucial conditions pertain to the sociopolitical institutions under which markets operate. Most European countries experienced, in this context, major changes. A variety of arrangements affecting the operation of labor markets were introduced by European governments. These arrangements include measures of employment protection, liberalizations of unemployment support, increasing payroll and social security taxes imposed on employers and similar measures. Employment protection, represented among other procedures by increased compensation payments upon dismissal, raised the expected real cost of employment relative to the real return expected from employment. The same result holds for increased payroll taxes and other obligations associated with employment. It follows that either net real wages fall or employment declines relatively. These adjustments are unavoidable. As it happened, the resulting adjustments occurred in most nations dominantly in the employment-unemployment dimension. No single step or single measure introduced in this evolution involved any dramatic or crucial changes. But their cumulative effect over time did change the operation of the economy. Whatever the motivation and intention of the political decisions shaping this evolution they produced a stagnant labor market. Vast human resources are poorly used, and nations are significantly poorer than they otherwise would be. The petrification of labor markets caused by a long sequence of political decisions also endangers the future course of Western societies. Technological innovations—a necessary condition of rising wealth in our history—will increasingly evolve as a social threat in the context of petrified markets. Under the circumstances the resulting political decisions tend to obstruct the wealth-creating process even further. Proposals to lower the working time per week to 35 or 25 hours illustrate this point.

The evolution of labor markets, however important, offers just one strand to our theme about the wealth and poverty of nations. The government's fiscal operations deserve some attention in this respect.

In an investigation of the consequences of taxes and government expenditures on the pattern of resource allocation, it was estimated that the general welfare of the United Kingdom was lowered by about 8 percent. Thus taxes and most particularly expenditure patterns severely distorted the allocation of resources.

A reference also should be made to the rising tide of protectionism. The range of protectionist schemes has expanded and evolved in imaginatively subtle complexity. The rationale for and the welfare implications of protectionism have barely changed over the centuries. Protectionist policies continue to redistribute wealth, benefiting special groups at the expense of the rest of society. The highly visible benefit of the favored group in conjunction with the widely dispersed—and thus less visible—social costs misleads politicians and publicists into believing that protectionist institutions increase general wealth. The redistribution produced by such policies is typically associated with a comparative social impoverishment. The ideology and illusion of protectionism as a necessary condition for fostering wealth creation permeates France, South America, and influential groups in many other nations.

Conclusion

History demonstrates, analogously to biological experimentation via mutation, an endless experimentation with social organizations and associated cultural terms, attitudes, and values. These organizations yield very different survival characteristics in competition with other social groups. They also determine the long-run chances of rising wealth or entrenched poverty and disease. The intentional, and frequently unintentional, evolution of the sociopolitical institutions decisively determines, ultimately, the conditions of poverty or wealth creation irrespective of the evolution's motivating conceptions. The role of political structure, represented by the state and its apparatus, thus deserves careful and intense attention in this context. The illusion is widely held that the state produces wealth, and, more particularly, that little wealth will be created without the detailed and controlling intervention of the state.

Another illusion dominating our time and well represented among the Christian churches holds that government intervention addressed to massive redistribution cannot affect the incentives guiding the social process of wealth creation. The translation of moral fervor "to help the poor" into institutional arrangements that lower opportunities and entrench poverty is one of the saddest ironies of our age. It is a small step from such beliefs to the view, implicit in the legal thesis of "tax-expenditures," that all wealth really belongs to the state and is conditionally and revocably leased to individual agents.

Our discussion reveals the nature of the misconception asserting a *direct* productivity of the state. The state is not a producer of wealth. It shapes *conditions* that encourage the creation of wealth. But it also frequently represents political institutions that impede expanding welfare. The state can, and frequently does, obstruct the wealth-creating process and contributes to sustained poverty. Its wealth-impeding activities yield an economic rent to a small group with access to the sociopolitical institutions. The emerging social organization of Western societies will thus determine whether a nation accumulates wealth or persists in poverty.

References

Bauer, Peter T. "Western Guilt and Third World Poverty." In *The First World & the Third World.* Edited by Karl Brunner. Rochester, N.Y.: University of Rochester Policy Center Publications, 1978.

Buchanan, James M. *The Limits of Liberty: Between Anarchy and Leviathan.* Chicago: University of Chicago Press, 1977.

Hayek, F. A. *Law, Legislation and Liberty.* Reprint (3 vols. in 1). London: Routledge & Kegal Paul, 1982.

Meckling, William H., and Jensen, Michael C. "A Positive Analysis of Rights Systems." Working paper, University of Rochester, Graduate School of Management, 1980.

Nozick, Robert. *Anarchy, State, and Utopia.* New York: Basic Books, 1974.

PROPERTY RIGHTS, LAND REFORM, AND ECONOMIC WELL-BEING

Victor A. Canto

Introduction

The basic premise of this paper is the belief that ownership enhances economic performance. Aside from token reference to political stability, little or no reference is usually made to the role of property rights in economic development.[1] Nevertheless some economists have recognized that the assignment of property rights is an important step in promoting a higher standard of living for any country's citizenry.

By pursuing different economic policies, governments can alter the allocation of property rights and, in so doing, can also alter the economic and political environment of a given country. At one extreme, governments can choose to promote an environment conducive to the use and exchange of private property. At the other extreme, they can promote the egalitarian distribution of property rights, including the enforcement of communal ownership.

Two basic political options are available to these governments. One option can be characterized as a political laissez-faire, that is, a process in which individuals express their choices through democratic elections. The other option is one in which the government curtails liberties and makes many or all the choices for the people.

A widely held view in Latin America postulates a trade-off between political freedom and economic prosperity. A cursory examination of many Latin American countries seems to support this view. In general, most of the Latin American countries where civil liberties are

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Associate Professor of Finance and Business Economics at the Graduate School of Business, University of Southern California. He wishes to thank David Boaz, James Dorn, and Charles Kadlec for their comments on earlier versions of this paper.

[1] "Property rights" is defined here as rights associated with uses of economic goods. These include rights to make decisions as to the use and transferability of the goods in question.

most rigorously observed are countries with a redistributive economic system. Examples include Jamaica during the Michael Manley administration, and Chile during the Eduardo Frei and Salvador Allende administrations. Although not technically part of Latin America, Puerto Rico is another example. Similarly, many of the high-growth countries in the early and middle 1970s did not have true democracy—for example, Chile under the military junta and Panama under Omar Torrijos—or, if they did, it was an autocratic government as the Dominican Republic had during Joaquin Balaguer's presidency.

The assertion that Latin American countries must make a choice between political freedom and economic freedom is not only unnecessary but misleading. If anything, political freedom is an important condition for economic prosperity, and vice versa. In the absence of economic freedom, political freedom ultimately is undermined. Symmetrically, without political freedom, economic freedom ultimately is lost to the pervasive power of the government. Both conditions foment change—either toward more freedom or toward curtailment of civil and economic liberties. In some cases the change occurs within a democratic system. Countries may move to the left, as was the case of Chile under Allende, or to the right, as in Jamaica under Edward Seaga. At other times the change may occur abruptly with bloodshed, as happened in Chile under Augusto Pinochet and in Nicaragua under the Sandinistas.

Political freedom and economic freedom complement one another. The role of personal incentives provides a strong argument against government intervention and an equally strong argument for individual freedom. Economic growth will occur in the property rights structure if a country makes socially productive activity worthwhile.

The possibility of changing ownership in private firms leads to specialization in the ownership where owners find their highest return. The possibility of changing ownership leads to a stronger association between efforts and rewards.

The Role of Government

Two important aspects of the relationship between government policies and property rights are (1) the effects of government policy in the allocation of property rights, and (2) the effects of property rights on political stability.

Historically, governments in the Western world have provided protection for and enforcement of property rights. In such a world, rational economic agents respond to incentives, and resources are allocated to their highest-valued uses. Similar arguments apply to the allocation of factors of production across countries. That is, in the absence of any friction, such as transportation cost and/or taxes, differences in prices, wages, and rates of return on capital across domestic and national boundaries are eliminated either through trade in goods or by factor movements (that is, migration and/or foreign investment).[2]

Through tax and expenditure policies, governments can alter the assignment of property rights. Such changes, however, affect incentives to invest in productive activities within their political jurisdiction, as well as the choice of investments.

It has long been recognized that the economic unit on which taxes are levied is not necessarily the unit that ultimately will incur the full cost of the tax. The final incidence or burden of a tax is dictated by price sensitivity, or elasticities of supply and demand for a good or factor on which the tax has been levied.[3] In either case, the more elastic (price sensitive) the supply, the lower the ability of the government to tax that factor of production. Similarly the lower the elasticity, the higher the government's ability to tax the factor of production.[4]

Supply elasticity may be large for one of two reasons. First, the factor of production is highly mobile across national boundaries; any attempt to lower its after-tax return below the return earned in the rest of the world will cause the factor to leave the country. Second, the factor of production has the ability to switch easily to other nontaxed activities (the underground economy, leisure, etc.); any attempt to lower its after-tax return will cause the factor to cease performing the taxed activity.

A strong argument can be made that in Latin America the least mobile factors are unimproved land and unskilled labor. The most mobile factors are capital and skilled labor. The above analysis suggests that broadly speaking, any attempt to tax mobile factors will result in those factors switching to nontaxed activities or migrating from the country.

The inability to tax a mobile factor directly is generally recognized by governments throughout the developing world. Evidence of this exists in the various forms of tax-holiday programs, free zones, or other methods of exempting mobile capital from domestic taxes.

---

[2] The conditions under which trade in goods is sufficient to equalize factor prices are well documented in the economic literature. See, for example, Samuelson (1948). For the effect of factor mobility on the equalization of factor prices, see Mundell (1957).

[3] On this issue, see the classic paper by Harberger (1962).

[4] For a discussion of the optimal tax structure along the lines discussed in the text, see Canto and Joines (1983).

Nontraditional exports also are granted partial or total tax exemptions. By contrast the consumption of luxury goods (air conditioning, automobiles, etc.), as well as traditional exports (usually agricultural products), is heavily taxed.

Governments, however, typically fail to recognize that any attempt to tax indirectly those factors of production that are mobile ultimately will be borne by the immobile factors. A tax on traditional exports, for example, will induce a reduction in the amount of capital utilized in the agricultural sector. Furthermore, since capital is mobile, the after-tax return earned in agriculture must be the same as in all other activities either inside or outside the country. The burden is borne by those fixed factors. Farm wages and the return to land will decline.[5]

This analysis suggests that depressed rural wages and the apparently inefficient production techniques used by landowners are attributable neither to exploitation nor lack of entrepreneurial ability. Rather, they are the direct result of the government's overall domestic policies. Over time this effect becomes even more pronounced. The tax burden reduces the landowners' incentives to improve or to keep up the quality of the land (less fertilizer is used, etc.). As this occurs, productivity and the salaries of farm workers decline.

The cost of political instability also is borne disproportionately by the least mobile factors—the land and unskilled labor. From an economic perspective, political instability raises the risk of reassignment of property rights. The result is not only reduced incentives to engage in productive activity but also the creation of incentives to devote resources to uses intended to minimize the potential economic losses arising from the loss of property rights.

The notion of political risk in the economics literature generally has been confined to risk analysis of countries. Such analysis usually deals with the possibility of an investment in a politically unstable country being expropriated with no compensation. Traditionally the way foreign investors have dealt with such risk has been by increasing the required rate of return on their investment. The end result has been lower foreign investment and, as a consequence, lower employment levels and wages within the country in question.

Outright expropriation with no compensation is rare. More common is inadequate compensation or incomplete expropriation. Examples of incomplete expropriation abound, in such forms as exchange controls and taxes. An example of inadequate compensation would be land reform where landowners were not paid a fair market price for their land.[6]

---

[5] A formal analysis of these propositions is developed by Canto (1983).

In the case of unskilled labor and land, there are striking similarities between inadequate compensation and incomplete expropriation. Because the incidence of direct and indirect taxes ultimately is borne by the fixed factors of production, the after-tax return to unskilled labor and land will decline as taxes rise anywhere in the economy.

A similar phenomenon may be true of land reform. In the case of partial expropriation through inadequate compensation, landowners clearly will suffer a loss of wealth. The effect of land reform on the wealth of rural workers, however, is not at all clear. Initially, giving the expropriated land to the workers increases their wealth. But to the extent that the political risk of the country increases, the required return on capital will increase. Hence the sum of workers' wages and the return to land will decline. In addition, to the extent that rural workers and land remain the inelastic factors of supply, they will ultimately suffer the burden of any increase in taxes to fund land reform and associated credit policies.[7]

Land reform is merely another attempt to redistribute wealth by taking from the rich and giving to the poor. However, economic forces that are beyond the control of any small country typically frustrate such attempts. The experience of land reform in Mexico—the *Ejido*—provides a clear-cut illustration of the fallacy of attempting to improve the economic well-being of the poor through land reform.

Mexican Land Reform: *Ejido* versus Private Sector Farms

Mexican land reform was introduced by legislation during the years 1915 to 1917. The unit of land that is distributed to rural workers is called an *ejido*. Each *ejido* consists of two kinds of land: noncultivatable land, which is essentially an open range used collectively; and cultivatable land, which may be farmed either individually or collectively by members of the *ejidos*.

The *ejidatarios* receive a use title that may be passed on to their heirs. Otherwise, land tenure is a nonnegotiable item. The security of the land is determined only by the *ejido* holder's presence in the community to defend his title. *Ejidatarios* have no right to sell, rent, lease, or mortgage their allotment of land. This restriction is intended to be a safeguard against the upper classes reacquiring large land holdings through foreclosure.

---

[6] The analysis presented assumes that current landowners are the rightful owners of the land. If they have acquired the land by illicit means, however, our analysis suggests that as a guarantor of property rights, government should develop the proper mechanism for the rightful owners to reacquire their land.

[7] See Canto (1983).

In general, arguments for land reform have been associated with political objectives. Increasingly, however, a focus on economic analysis is being used to justify land reform. The bulk of these economic analyses compare the productivity of the land-reform sector with that of the nonreformed, private sector. Several recent studies, including those by Dovring (1970), Muller (1970), and Nguyen and Martinez (1979), have provided evidence that contradicts the view that the land-reform sector is inefficient.[8]

The Nguyen and Martinez study's statistical analysis yields some interesting results:

- Both *ejidal* and private-sector farms exhibit positive rates of returns.
- The mean marginal yield for *ejidal* farmers is higher than for private-sector farmers.
- *Ejidal* farmers have a lower elasticity of returns to total expenditures than do private-sector farmers. The authors attribute this, in part, to the higher average marginal yields of *ejidal* farmers.

The empirical results reported by Nguyen and Martinez provide convincing evidence that the productivity of *ejidal* farms is as good if not better than that of private sector farms. These results, however, are somewhat misleading.

The higher marginal return of *ejidal* farmers is easily explained in terms of credit availability. *Ejidatarios* do not hold title to their land and cannot pledge their land as collateral. The only collateral they can use is their future crops. This limitation increases the risk of lending to *ejidatarios*. Consequently, either *ejidatarios* will be charged a higher interest rate or only the most productive—or lower-risk—*ejidatarios* will obtain loans. Either alternative explains the higher marginal yield of the *ejidos*.

In addition, the Nguyen and Martinez analysis of the productivity of *ejidal* farms overestimates their profitability. Nguyen and Martinez do not incorporate into their cost calculation the value of the time spent by *ejidatarios* and their relatives working on the *ejidos*. One common argument made to minimize this omission is that unemployment in the developing nations is so great that for all practical purposes, labor can be considered a free resource.[9] This argument, however, does not change the fact that private sector farmers must pay wages to their workers. Thus, given that in some sense *ejidatarios* or their relatives have the option of working on private sector farms, their work time should be valued at the going wage rate. Moreover, the *ejidal* sector has a considerably higher labor/land ratio without having a higher yield per unit of land. It necessarily follows that once the value of the *ejidatarios* is factored into the total cost, the *ejidos* are substantially less profitable than private-sector farms. Furthermore, the higher labor/land ratio and equal yield of *ejido* farms suggest that people in the *ejidal* sector have a lower standard of living than those in the private sector.

---

[8] An interesting discussion on the *ejido* system is found in the cited publications.

[9] This assumption is commonly used in the evaluation of public projects in developing countries. For a discussion of the fallacy implicit in such an assumption, see "Social Opportunity Cost of Labor" in Harberger (1974).

Although in principle *ejido* land can neither be rented nor sold, and is transferable only to the user's heirs, in practice much of it has been rented for many years to outsiders or has been sold in violation of the Mexican constitution. The agrarian authorities may have been unable or unwilling to stop this practice.[10] One of the many ways the *ejido* land transferability limitation is circumvented is through sharecropping arrangements. The sharecroppers are basically small businessmen who maintain a labor force of their own and, in many cases, provide financing for the crops. As part of the basic arrangement, they lend the *ejidatarios* a sum of money to sharecrop the land. The "loan" does not bear interest and is not repayable until such time as the landowner decides to discontinue the sharecropping arrangement. As a rule of thumb, the sharecroppers keep 50 percent of the profits, which gives them an equity position in the land. In short, the sharecropping arrangement is one way to circumvent the sale, lease, and rental restrictions on *ejido* property.

The sharecropping arrangements also defeat the basic objectives of the *ejido* system: to break up monopolistic control of the land and to create a symmetrical social and economic relationship between peasant and landholder. Ironically, now that the *ejidatarios* (who were formerly landless peasants) own the land, they have become dependent on landless sharecroppers. To the extent that sharecroppers, who are a mobile factor, hold control over the *ejidatarios,* they will capture a larger share of the profits than the *ejidatarios*, thereby defeating the basic objectives of agrarian reform.

In short, after 60 years, land reform in Mexico has had little or no impact on the distribution of income or wealth among its citizens. Moreover, while Mexican land reform has done well in terms of productivity (output per acre of land) and employment objectives, it has done little to improve the standard of living of the country's formerly landless peasants.

---

[10] On this issue, see Stavenhagen (1973).

Land Reform in El Salvador

El Salvador's land reform is viewed by many as a direct result of U.S. foreign policy during the Carter administration, and as continued through the Reagan administration.[11] The political motives for the land reform are fairly well established. In the words of Secretary of State George Shultz, "Promotion of land reform stands at the very heart of El Salvador's effort to encourage social equity, political stability and economic development."[12] The land reform consisted of three phases: phase I was to nationalize all farms in excess of 1,250 acres, and phase II was to expropriate farms of between 250 and 1,250 acres. For political reasons, phase II has not been implemented.[13] Under phase III, El Salvador's bank must give priority to the newly financed cooperatives under the land reform. The big farms held by the oligarchy—194 in all, accounting for 11 percent of the land—were nationalized and converted into 317 cooperatives with some 31,000 members.

In 1982 proponents of the land-reform program cited a study that found first-year production of the land taken under phase I to have declined by perhaps 10 percent. The study also noted that second-year production had rebounded and was expected to increase by 10 to 20 percent. Alternatively stated, proponents of land reform said that the reform had no adverse effect on the productivity of the expropriated farms. This information was then used as evidence that reform had not caused severe economic damage and had solved many of the grievances of the previously landless Salvadoran peasants.[14]

If the Mexican experience is any guide, a strong case can be made that productivity figures quoted in the El Salvador study do not provide evidence of an improvement in the distribution of income. One only needs to be reminded of the fact that the "benefactors" of land reform use a considerably higher labor/land ratio without having a higher yield than private sector farmers. The evidence available three years after the March 1980 land reform indicates that except for corn and bean production, the production on nationalized farms has fallen: Coffee production declined by 50 percent, sugarcane by 43 percent, and cotton by 55 percent.[15] These results were contrary to the initial expectations of proponents of the reform.

---

[11] A brief account of U.S. policy toward El Salvador is presented in "Democracy When It Suits" (1982) and "Salvador's Fibrillating 'Heart' " (1984), as well as in Alfamiano (1982).

[12] See "Salvador's Fibrillating 'Heart' " (1984). For a very persuasive counterargument to Shultz's views, see Ayau (1984).

[13] For a discussion of the measures adopted to curb land reform, see Sieb (1982).

[14] See, for example, Prosterman (1982).

The decline in agricultural output on nationalized farms is attributable in part to El Salvador's war effort. The U.S. embassy in El Salvador estimates that of the 325 farms turned into cooperatives under the land reform, 40 are in the war zone, 95 are holding their own, 95 have serious problems that may be resolved, and 95 are beyond hope. This indicates that even after accounting for misfortune and the fall in international prices, production undeniably has declined. It necessarily follows that the standard of living of peasants in the land-reform sector will be below that of private sector farmers. Whether their well-being increases compared to their previous status will depend upon whether titles to the land are transferable through sale, rent, or lease.

The combination of land reform, depressed international prices, and civil war has caused many of the employees-turned-land holders to look for employment elsewhere. At many nationalized farms the wage rate has been slashed by more than the decline in international prices.

If land titles could be sold, then those individuals who acquired land from the 1980 reform but who lacked entrepreneurial capacity and sufficient credit-worthiness would sell their titles, and land would flow to more productive uses. In addition to their wages as employees, these individuals would receive a one-time wealth transfer from the sale of their land. On the other hand, those peasants with entrepreneurial capacity and sufficient credit would have an incentive to retain titles to their land. And, to the extent that these individuals make up only a small fraction of the population, land tenure would quickly revert to the pre-reform ways. The option to sell one's property title, however, appears to have been already ruled out by Decree 207, which makes it illegal for new owners to sell or rent their land.

If this decree remains in place, it is likely that Salvadoran ingenuity, like that of the Mexican *ejidatarios*, will find a way to circumvent the restrictions on the sale of land. Sharecropping arrangements can be expected to flourish as peasants lacking entrepreneurial capacity and/or credit enter into sharecropping agreements. Again, the new landowners will be better off to the extent that they receive one-time, interest-free loans to enter into sharecropping agreements.

---

[15] Corn and bean production were maintained because of the larger acreage devoted to these staples. See "Salvador's Fibrillating 'Heart' " (1984).

Such an arrangement, however, will ultimately reduce the efficiency of the land-reform sector. Because the land is nontransferable, the land-reform sector will be a higher credit risk. Productivity of the land and the well-being of the farmers may be reduced further if the land-reform program does not take into consideration the fact that some plots of land in El Salvador are often left fallow for one to four years. This consideration alone suggests that the plots will have to be two to five times larger than at present to support the farmers.

Several other economic considerations further reduce the attractiveness of the Salvadoran land-reform program. The reform will impose an enormous and costly administrative burden on the government. Also, to the extent that the farmers truly pose a high risk to lenders and that the sources of funds dry up, the government will have to develop special credit programs, thereby further distorting the economy. In addition, to the extent that funds are fungible, there is no guarantee that the funds lent will be put to the use for which they were ostensibly borrowed.

Evidence of poor production in land-reform farms is well documented.[16] Consider, for example, the poor performance of the cooperatives. A recent U.S. Agency for International Development (AID) report says that cooperatives have a larger-than-needed labor force. This evidence clearly indicates that phase III has not improved management.[17]

Another consequence of the land reform is that it reassigns property rights, which in turn creates an environment of economic uncertainty. This not only results in reduced incentives to engage in productive activities but also creates incentives to devote resources to uses intended to minimize potential economic losses from arbitrary property rights reassignment. Such activity would include attempts by the upper classes to relocate a portion of their wealth outside the jurisdiction of the Salvadoran government. These symptoms are already evident in El Salvador. As the business community attempts to take capital out of the country, these outflows create substantial balance-of-payment pressures. Rather than pledging to guarantee the property rights of business owners, the Salvadoran government has responded to the outflow by nationalizing banks and export companies, thereby further increasing the country's political risk.

Additional side effects of the land reform include capital losses suffered by previous landowners paid in government bonds for their confiscated farms. These bonds are trading at 50 percent of their face value. Many farmers, as a result of the uncertainty surrounding phase II of the land reform, did not bother to plant any crops on their farms because of the risk of expropriation.[18]

---

[16] On these issues, see Schuster (1982a, 1982b). More recent information points to the continuing decline in production; see, for example, "Salvador's Fibrillating 'Heart' " (1984).

[17] See "Salvador's Fibrillating 'Heart' " (1984).

The net effect of this increase in uncertainty from the land reform and its accompanying policies has been to increase the return required by potential foreign investors. In other words the availability of private capital, an already scarce resource in El Salvador, will more than likely decrease in a significant manner. With less capital to work with, the overall efficiency of the economy is likely to decrease, thereby reducing the overall well-being of the Salvadoran people.

Other available information on El Salvador suggests that many of Mexico's experiences with the *ejido* system will be repeated in El Salvador if land reform as advocated by the United States is fully implemented.

Failure of the Redistributive Model

In light of the experience of other Latin American countries, the available data for El Salvador suggest that Salvadoran land reform may radically alter property rights and increase political risks in that country.[19]

The net effect of these policies will reduce direct foreign investment and result in significant capital outflows. Judging from recent experience, the expected reaction of the government will be to infringe even further on current property rights, which in turn will only reduce the efficiency and overall level of economic activity in El Salvador.

The productivity of the land-reform sector (output per acre of land) may be as high as that of the private sector. It is highly unlikely, however, that land reform will lead either to a more equitable distribution of wealth and income or to a generalized improvement in the standard of living of the country's poor. Land and unskilled peasants are the least mobile factors of production. As such, the burden of financing government expenditures ultimately will fall on them, thereby reducing their after-tax income.

The problems of insufficient income among the rural peasants in El Salvador, as well as in other developing nations, are fundamentally economic problems. As such, they require an economic, not a political, solution. Using the Latin American experience as a guide, numerous examples can be found in which politically motivated redistributive policies have failed. Examples include those involving Chile during the administrations of Eduardo Frei (1964–70) and Salvador Allende (1970–73).

---

[18] See Schuster (1982a, 1982b).

[19] See Ayau (1983) for a persuasive argument that much of Guatemala's political unrest can be traced to its 1952 land reform legislation.

The objectives of the Frei administration included the nationalization of major industrial companies and utilities. In its first year, the administration instituted a wealth tax to finance these programs. This led to substantial increase in government expenditures. The Frei administration relied mainly on internal deficit financing and introduced a "crawling peg" exchange rate system. Continued currency devaluation averaged 30 percent per year.

The Allende government may be viewed as having carried to their logical extremes many of the reforms initiated by the Frei administration and now advocated by the U.S. government for implementation in El Salvador. Under the Allende administration, property rights were abrogated throughout the economy. Major industries and utilities were nationalized, a comprehensive welfare system was enacted, and price controls on food products were instituted. Unemployment was fought by increasing employment in government-owned mines. The goal was full employment, income redistribution, and price stabilization. Real wages and social security pensions were increased by 32 percent. The tax system also was made more progressive, wealth taxes were eliminated for the lower third of the population, and the income tax exemption was raised to two times the *Sueldo Vital* (the minimum guaranteed salary), thereby essentially excluding 20 percent of the population from paying the tax.

All this led to a dramatic increase in government expenditures as a share of gross domestic product. By 1973 this ratio had reached an all-time high of 43 percent. The Allende government lost most of its international credit, owing in part to political pressure from the United States. Money creation became the only viable route to finance public expenditures. Money supply growth accelerated from 35 percent in 1965 to 345 percent in 1973. In the same period, inflation spiraled upward from 29.3 percent to 316 percent.

As a result of rigid price controls, black market activities flourished. The government's nationalization policy contributed to substantial capital outflows. These factors, combined with galloping inflation and work stoppages, led to the collapse of Chile's government. The economic and political crises produced by the failure of these policies paved the way for the 1973 coup d'etat, when Allende was overthrown by the military.

The case of Jamaica, like that of Chile, is an example of the failure of the redistributive model. The Jamaican experience, however, illustrates that the changes from one system to another can occur within a democratic process. The election of Edward Seaga as prime minister in 1980 was a repudiation of previous interventionist economic policies in Jamaica. As indicated in Table 1, the Jamaican inflation rate was in double digits in all but one of the years of the Manley administration, and the growth rate of real output was negative in every year except 1973. All this occurred while government intervention increased substantially. Government expenditures (consumption) as a share of gross national product increased from 13.9 percent in 1972 to 23.2 percent in 1980. During this same period, direct foreign investment declined, becoming negative in 1975 through 1979.

Conclusion: Finding a Solution for Latin America

The experiences of Jamaica and Chile suggest that as economic policies failed, they precipitated a change to a different economic model. In the case of Jamaica, these changes occurred within a democratic process when Michael Manley's administration was voted out of power. In the case of Chile, the redistributionist policies, in addition to precipitating economic changes, resulted also in changing the political process: The democratically elected Allende administration was overthrown by the military.

At the other end of the spectrum lie those governments that repress civil liberties while pursuing economic liberties. Although the economic costs of such a system may be small, the costs in human life and civil liberties, which are difficult to quantify, are substantial. As a result, authoritarian governments also are inherently unstable. Chile, under Augusto Pinochet, appears to be an example of economic liberalism and political suppression. Recent events seem to signal an end to the Chilean economic miracle. It remains to be seen whether the economic failures of the junta will precipitate a political change as well. In short, the Latin American experience suggests that the combination either of economic liberalism and political suppression or of political liberalism and economic repression results in an unstable system.[20]

---

[20] Our analysis does not explicitly consider the case in which both economic and civil liberties are curtailed. In fact, in Latin America, one can find examples of both totalitarian and authoritarian governments that appear to have endured the test of time (for example, Castro in Cuba and Stroessner in Paraguay). However, it is also fair to point out that previous experiences there have been authoritarian governments that have lasted longer (for example, Trujillo in the Dominican Republic and the Somozas in Nicaragua). In both cases the regimes ended abruptly, but the two countries adopted vastly different systems: The Dominican Republic adopted a democratic process while Nicaragua opted for a totalitarian government.

TABLE 1

SELECTED ECONOMIC STATISTICS ON THE JAMAICAN ECONOMY DURING THE MANLEY ADMINISTRATION

Year	Inflation Rate	Real Output Growth Rate	Government Expenditures as % of GNP	Direct Foreign Investment[a]	Merchandise Trade Balance[a]
1972	18.0	−2.6	13.9	97.8	−151.90
1973	21.8	1.3	16.4	75.1	−178.30
1974	16.0	−2.3	17.3	23.3	−59.10
1975	9.2	−6.9	18.5	−1.8	−161.00
1976	10.8	−2.6	21.4	−0.6	−135.10
1977	29.9	−0.3	21.5	−7.0	70.10
1978	25.5	−1.5	21.5	−26.7	80.20
1979	24.0	−5.5	21.2	−26.4	−64.30
1980	12.7	−15.5	23.2	27.6	−74.50

[a] Millions of U.S. dollars.

SOURCE: International Monetary Fund, *International Financial Statistics* (Washington, 1982).

Required Policies

The solution to the Salvadoran problems (as well as to those of other Latin American countries) requires a reorientation of economic policies. Elevating the standard of living of the poor requires a complete restructuring of El Salvador's economic policies:

- A first and necessary step is to reinstitute the government in its role as protector and enforcer of property rights by ending land reform, nationalization, and other associated policies. To succeed, new policies must be able to attract new capital, which in turn will increase the demand for the domestic factors of production.
- Second, these policies must also reduce the burden of the tax system on the traditional sectors of the economy.
- Third, the government must develop programs to increase the skill levels and hence the mobility of poor people.

The Role of the United States

The United States can contribute a great deal to the achievement of a desirable political and economic balance in Latin America. Politically, it should encourage continued progress in human freedom without directly interfering in internal political affairs. Economically, it can provide aid—although not direct assistance, which serves only to create economic dependency. The best aid the United States can provide is to allow the Caribbean and Central American countries access to U.S. markets.[21] This action, including a reduction in the current trade restrictions on a number of agricultural and labor-intensive manufactured products, would (1) provide higher revenues to the national governments which would enable them to implement their own development plans; and (2) increase the total price received by local producers, which would increase the profitability of the traditional export industries, and thus provide an incentive for capital to flow into these areas.

In the process, the level of employment and well-being of the poor of Central America would be improved. The United States, therefore, would contribute in a significant way to the strengthening of economic development and to the democratic process in El Salvador.

References

Alfamiano, E. "The Threat to Cut Off Aid to El Salvador." *Wall Street Journal*, 25 June 1982, p. 18.

---

[21] Krauss (1984) reaches a similar conclusion.

Ayau, M. "Land Reform Often Reaps a Crop of Troubles."*Wall Street Journal*, 23 September 1983, p. 31.

Ayau, M. "Land Reform: Wrong Lesson, Wrong Result." *Commentary* (June 1984), p. 9. Universidad Francisco Marroquin, Guatemala.

Canto, V. A. "The Incidence of Land Reform." Working paper. University of Southern California, 1983.

Canto, V. A., and Joines, D. H. "Taxation, Revenue and Welfare." *Economic Inquiry*, 21 July 1983, pp. 431–38.

"Democracy When It Suits" (editorial). *Wall Street Journal*, 10 June 1982, p. 30.

Dovring, F. "Land Reform and Productivity in Mexico." *Land Economics* 46 (August 1970): 264–74.

Harberger, Arnold C. "The Incidence of the Corporation Income Tax." *Journal of Political Economy* 70 (June 1962): 215–40. Reprinted in *Taxation and Welfare*, pp. 135–62. Edited by A.C. Harberger. Boston: Little, Brown and Co., 1974.

Harberger, A. C. *Project Evaluation: Collected Papers*. Chicago: Markham, 1974.

Krauss, M. "U.S. Objective in Central America: Promoting a Privatized Economy." *Detroit News*, 20 February 1984, p. 46.

Muller, M. "Changing Patterns of Agricultural Output and Productivity in the Private Land Reform Sectors in Mexico, 1940–60." *Economic Development and Cultural Change* 18 (January 1970): 252–66.

Mundell, R. A. "International Trade and Factor Mobility." *American Economic Review* 47 (June 1957): 321–35.

Nguyen, D., and Martinez, M. "The Effects of Land Reform on Agricultural Production, Employment and Income Distribution: A Statistical Study of the Mexican State 1959–69." *Economic Journal* 89 (September 1979): 624–35.

Prosterman, Roy L. "Salvador's Alternative to Revolution," letter to the editor. *Wall Street Journal*, 6 May 1982.

"Salvador's Fibrillating 'Heart' " (editorial). *Wall Street Journal*, 21 February 1984, p. 32.

Samuelson, P. A. "International Trade and the Equalization of Factor Prices." *Economic Journal* 58 (June 1948): 163–84.

Schuster, L. "Multimillion Dollar Washington Aid to El Salvador is Only a Stop Gap at Best." *Wall Street Journal*, 8 May 1982a, p. 30.

Schuster, L. "Agrarian Mire: Land Reform Proves Costly to El Salvador, Mixed Blessings to Poor." *Wall Street Journal* 1 September 1982b, p. 1.

Sieb, G. I. "Move by El Salvador to Curb Land Reform Put U.S. Plans for Aid in Some Jeopardy." *Wall Street Journal*, 21 May 1982.

Stavenhagen, R. "Land Reform and Institutional Alternatives in Agriculture: The Case of the Mexican *Ejido*." Occasional paper no. 73-9, Vienna Institute for Development, Vienna, Austria, 1973.

RESOURCES, PEOPLE, AND THE NEOMALTHUSIAN FALLACY

David Osterfeld

Introduction

"It isn't ignorance that is so bad," Will Rogers is reputed to have said, "but all those things we know that just ain't so." For the last two decades or so we have been fed a steady diet of "scientific" studies showing that the world is on the brink of disaster.

Of course catastrophism, in one form or another, is nothing new. It can be traced back at least to Thomas Malthus and David Ricardo nearly two centuries ago. But over the last two decades the catastrophist or Neomalthusian position has achieved dominance in the world of ideas. The new wave of Neomalthusian literature is so vast that no review can hope to be complete. But mention should be made of the studies by Paul Ehrlich (1968, 1969; and 1974 with Anne Ehrlich), Donella Meadows et al. (1972), Garrett Hardin (1968), Barry Commoner (1976), William and Paul Paddock (1967), Richard Barnet (1978, 1980), Nicholas Georgescu-Roegen (1971, 1977), Herman Daly (1974, 1979), and Lewis Perlman (1976), to name but a few. Neomalthusianism became the official position of the U.S. government with the publication of the *Global 2000 Report*, prepared in 1980 by the Council on Environmental Quality in conjunction with no less than 12 other government agencies. The most striking thing about these studies is not so much their bulk, but their scope. They have originated from or touched on a variety of disciplines including biology, physics, agronomy, demography, geology, political science, and economics. While there have been dissenters (Barnett and Morse 1963; Barnett 1979; Goeller and Weinberg 1978; Kahn 1976;

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Associate Professor of Political Science at Saint Joseph's College, Rensselaer, Indiana. He expresses his thanks to Jennifer Vohlken for her assistance in preparing this paper for publication.

Beckerman 1975; Nordhaus 1974; Simon 1980, 1981a, 1981b; and Simon and Kahn 1984), they have failed to have much impact. There is little doubt that Neomalthusianism is today's conventional wisdom on population growth, resources, pollution, and economic development. But much of this wisdom "just ain't so."

According to the Neomalthusians, the "unbridled growth" in population is creating "unmanageable population pressures" that are "holding back the advancement of the poor" and "will have catastrophic consequences" on a par with nuclear war. As population outstrips food supplies, the results will be "famine, riot, insurrection and war." A further consequence will be increasing competition for "dwindling supplies of the Earth's resources." This will cause "high inflation, massive unemployment, and economic and social stagnation or even decline" (McLaughlin 1982). The only solution lies in what Daly (1974, 1979) terms the "steady-state economy" and what Perlman (1976, pp. 179–81) calls "ecotopia" or "global equilibrium," that is, a world characterized by "zero growth in the stock of capital and zero growth in the stock of human population." Practically every link in this argument is weak.

Population Density, Population Growth, and Poverty

If the above argument were correct, one would expect to find a clear relationship between economic growth on the one hand and either population density or population growth on the other. The foregoing statement actually implies two quite different relationships between population density and per capita income. One is that the larger the population the greater the strain on physical resources and, hence, the lower per capita income. If correct, per capita income would decline as population densities increased so that the most densely populated countries would have the lowest per capita incomes. A more sophisticated version suggests that there is an optimum population density below which economies of scale cannot be exploited and above which there would exist crowding and resource drain. Thus, per capita income would decline as one moved away from the optimum. If correct, one would expect to find a curvilinear relationship between population density and per capita income.

Neither relationship holds. In fact, what is clearly evident when the data are examined systematically is the lack of any relationship between population density and per capita income. There are both rich and poor countries having high population densities; there are rich and poor countries with low population densities; and there are rich and poor countries between these extremes. For example, "overpopulated" Bangladesh, with 617 people per square kilometer (ppsk), has a per capita income of $90 per year. But Hong Kong has a ppsk of 5,000. Yet, it has a per capita income of $3,760. And Singapore has a ppsk of 2,400 coupled with a per capita income of $3,830. On the other extreme there are "underpopulated" countries such as Mali with a density of only 5.5 ppsk and a per capita income of only $140. Yet Canada, with a per capita income of $9,640 has a density of only 2.4 ppsk. Australia has a density of 1.9 ppsk coupled with a per capita income of $9,120. There are also rich and poor countries between these extremes. France has a density of 97 and a per capita income of $9,950. The People's Republic of China, however, has a density of 101.5 and a per capita income of $260, while Uganda, with a density of 54, has a per capita income of $290. Finally, the United States has a population density of 24 and a per capita income of $10,640. If these World Bank (1981) figures indicate anything, it is that per capita income is not determined by population density.

History does show that there is a relationship between population growth and economic progress, but the relationship is the reverse of that expected from conventional wisdom. The Industrial Revolution was a period of sustained, rapid economic and social advance without parallel. It was also a period of unprecedented population growth. In a study of the demographic aspects of the Industrial Revolution, Mabel Buer (1968) found that prior to 1750 "the proportion of those born in London dying before five years of age" was 74.9 percent. By 1820 it had fallen to 32 percent. Although probably lower in rural than in urban areas such as London, the significant point is the dramatic, across-the-board decline in infant mortality in both rural and urban areas in Europe as well as England. "The fall in the infant mortality rate," says Buer, "was common to the whole of northwestern Europe. . . . For Europe, deChateauneuf estimated that in 1780 about half of the children born . . . died before reaching the age of 10 years. In 1825 the corresponding figure was a little more than one-third" (pp. 30–31).

Consequently, in the 70 years from 1750 to 1820, the population of England doubled, going from 6 to 12 million. There is little doubt that the standard of living of these 12 million people in 1820 was far higher than that of the 6 million in 1750 (Ashton 1960, pp. 3–6). One finds similar patterns not only in 18th- and 19th-century Europe and America, where the standard of living increased dramatically even with a quadrupling of the population since 1850, but in several Third World countries as well. Malaysia, Singapore, Hong Kong, Brazil, South Korea, and Taiwan are cases in point. All have experienced sustained economic development together with rapid population growth.

It is significant that while England was "overpopulated" with a population of six million people in 1750, today it supports a population in excess of 50 million at a standard of living many times higher than that of 1750.[1] "Overpopulation," therefore, is relative. It depends on the state of technology, political institutions, and, most important, social mores and the personal attributes of the population. A culture that encourages hard work, thrift, and innovation will be far more productive and able to support a much larger population at a higher standard of living than a culture in which such traits are discouraged.

In fact, a sparse population may impede economic progress. There is a tendency when discussing population increases to look only at the number of mouths to be fed. But people are producers as well as consumers. For every mouth to be fed there is a pair of hands that can work and a brain that can think. Thus, although population growth stimulates demand, it also supplies the labor to meed that demand. The number of jobs is not fixed. There is therefore no reason that population growth in itself should cause unemployment. This was not the case in the West or in such densely populated non-Western countries as Hong Kong and Singapore. In fact, while it is one of the most densely populated countries in the world, Singapore continues to attract immigrants from its less densely populated neighbors.

Population growth increases both aggregate output and aggregate demand and thus expands markets. Because this expansion increases specialization and facilitates economies of scale, productivity tends to increase at a faster rate than population. It is not surprising that, as Bauer (1981, p. 49) points out, "contemporary famines and food shortages occur mostly in sparsely population subsistence economies with abundant land" such as Ethiopia, Uganda, Zaire, and Tanzania. A sparse population means that there is little specialization and hence low productivity. It also means that modern communications and transportation facilities are all but nonexistent. The consequence was graphically illustrated by the observation of a Red Cross officer during the famine in the Sahel in West Africa in the 1970s: "Sure, the food is pouring in, but how the hell are we going to get it to the people who need it? There isn't a tarred road within a thousand miles of Juba" (Simon 1981b, pp. 61–62). It is notable, says Bauer (1981, p. 49), "that no famines are reported from such densely populated regions of the less developed world as Taiwan, Hong Kong, Singapore, Malaysia and cash crop producing areas of West Africa."

---

[1] There is little doubt that England was overpopulated relative to the technology of the day. McEvedy and Jones (1978, p. 42) write: "[T]he population figure of around 3.75 million (in 1300) seems to have been well over the optimum for the agricultural technology of the time, for, as more and more marginal land was brought into use, both productivity and the standard of living fell. Since the norm was little better than subsistence, the nutritive state of the population declined dangerously. By 1300 the population was having difficulty maintaining itself and before the bubonic plague had ever hit England the stage was set for disaster." Similar conclusions were reached for Europe (p. 24).

There is another very important reason for the high productivity of many densely populated areas. Population density facilitates the exchange of ideas and leads to technological innovations and scientific discoveries, all of which increase productivity. An oppressive and stultifying environment retards the exchange of ideas, regardless of population density.

Neomalthusian Predictions: A Scorecard

The Neomalthusians base their predictions of imminent catastrophe on the notion that the supply of resources is finite. Population growth, they argue, threatens productivity by taxing the world's "diminishing resources." The more oil, copper, or natural gas we use, the less there is left, and the more quickly we consume them the sooner we will run out. In 1865 William Stanley Jevons predicted "the impossibility of a long continuance of progress" due to the impending exhaustion of coal supplies. Since then hardly a year has gone by when an "expert" or an "official body" has not solemnly predicted the "end of affluence" or the dawn of the "age of scarcity" due to the exhaustion of "nonrenewable" energy supplies or the depletion of key raw materials (see Table 1). Yet none of these predictions have materialized.

The predictions of today's Neomalthusians have proven no more accurate. Biologist Paul Ehrlich (1969) predicted for the 1970s: (1) "increasing poverty" and world hunger; (2) rapid deterioration of the "raw material situation"; (3) sustained "environmental deterioration," especially in the United States, resulting in "nearly 200,000 corpses" by 1973; (4) reduction in life expectancy to 42 years by 1980; (5) the "end of the ocean" by 1979; and (6) "the birth of the Midwestern desert," slated for the summer of 1973. It is a "pretty grim scenario," says Ehrlich, yet it is "based on projection of trends already appearing. We're a long way into it already." Each of these predictions will be examined here.

First, studies conducted by the United Nation's Food and Agricultural Organization (UNFAO) show that since 1948, the first year for which such data are available, the world food supply has consis-

TABLE 1

OIL PROPHECIES FROM OFFICIAL SOURCES AND THE CORRESPONDING REALITIES

Date	*U.S. Oil Production Rate (billion bbls./yr)*	*Prophecy*	*Reality*
1866	.005	Synthetics available if oil production should end (U.S. Revenue Commission).	In next 82 years the U.S. produced 37 billion bbls. with no need for synthetics.
1885	.02	Little or no chance for oil in California (U.S. Geological Survey).	8 billion bbls. produced in California since that date with important new findings in 1948.
1891	.05	Little or no chance for oil in Kansas or Texas (U.S. Geological Survey).	14 billion bbls. produced in these two states since 1891.
1908	.18	Maximum future supply of 22.5 billion bbls. (officials of Geological Survey).	35 billion bbls. produced since 1908, with 26.8 billion reserve proven and available on 1 January 1949.
1914	.27	Total future production only 5.7 billion bbls. (official of U.S. Bureau of Mines).	34 billion bbls. produced since 1914, or six times this prediction.
1920	.45	U.S. needs foreign oil and synthetics; peak domestic production almost reached (Director of U.S. Geological Survey).	1948 U.S. production in excess of U.S. consumption and more than four times 1920 output.

TABLE 1 (cont.)

Date	*U.S. Oil Production Rate (billion bbls./yr)*	*Prophecy*	*Reality*
1931	.85	Must import as much foreign oil as possible to save domestic supply (Secretary of the Interior).	During next 8 years imports were discouraged and 14 billion bbls. were found in the U.S.
1939	1.3	U.S. oil supplies will last only 13 years (radio broadcasts by Interior Department).	New oil found since 1939 exceeds the 13 years' supply known at that time.
1947	1.9	Sufficient oil cannot be found in U.S. (Chief of Petroleum Division, State Department).	4.3 billion bbls. found in 1948, the largest volume in history and twice our consumption.
1949	2.0	End of U.S. oil supply almost in sight (Secretary of the Interior).	Petroleum industry demonstrated ability to increase U.S. production by more than a million bbls. daily in the next 5 years.

SOURCE: Kahn et al. (1976, pp. 94–95).

tently exceeded world population increases by about one percent per year. Such aggregates obscure as much as they reveal. For example, the UNFAO data (1977, pp. 3–7) show that while the increase in per capita food production in the developed countries averaged 1.4 percent per year for the 15-year period from 1961 to 1976, production in the developing countries declined from a yearly average of 0.7 percent in the 1960s to 0.3 percent between 1970 and 1976. Similarly, per capita caloric consumption as a "percent of requirements" increased in the developed countries from 124 in the period 1961–63 to 132 in 1972–74. That for the developing countries went from 89 in 1961–63 to 96 in 1972–74. While this represents a substantial improvement, it is noteworthy that most of the gain for the developing countries occurred during the 1960s. Before examining the reasons for the decline in the rate of improvement, a note of caution is in order: A decline in the *rate* of improvement must not be confused with a deterioration in absolute terms. The fact is that the average Third World individual was better fed in 1980 than in 1969. Put differently, the world food situation, *even for developing countries*, is improving, not deteriorating.

But why the slowdown in the early 1970s, especially for the developing countries? Two nearly simultaneous but unrelated occurrences are often advanced as explanations. One is meteorological, the other political. In 1971 and 1972 nearly worldwide bad weather resulted in widespread crop failures. The UNFAO (1977, p. 3) reports that 1972 was probably "the only time since the Second World War [that] world food production declined slightly in absolute terms from the previous year's level and not just in relation to population growth." This was also the time of the dramatic increase in oil prices, which, according to conventional wisdom, put the "green revolution" out of reach for many of the petroleum-poor developing countries. But there is no one-to-one correspondence between oil prices and farm costs. What has gone largely unnoticed is that after controlling for inflation, fertilizer prices did not increase at all in the 18-year period from 1967 to 1980 (Johnson 1984, p. 105). Thus, although meteorological factors had an adverse but short-term effect, the impact of oil prices was negligible. Neither adequately explains the declining rate of increase in agricultural outputs throughout much of the Third World during the 1970s.

Given the Neomalthusian concern (Meadows et al. 1972, pp. 46–54) about the scarcity of arable land, it should be noted that this was not a factor either. Only about 12 percent of the earth's land mass is cultivated on a regular basis. Much of the remaining 88 percent is arable, and the largest percentage of uncultivated arable land lies in the developing countries. Agricultural economist W. David Hopper (1976, pp. 137–38) writes that

> most of the developing countries are better endowed for agricultural progress than for any other kind of economic advance. The developing world lies largely on and between the tropics of Cancer and Capricorn. It is a belt of warm temperatures, of generally abundant (if often seasonal) rainfall and of ample, year-round solar energy to be converted into chemical energy for storage in plant and animal tissue.

Despite these ideal conditions, Hopper points out that in Africa only 22 percent of the arable land is actually cultivated. In South America it is a mere 11 percent. D. Gale Johnson (1984, pp. 95–97) argues that the availability of arable land is "largely irrelevant" because it "ignores the changes that can be made to improve the productive capacity of that land and to increase significantly the stability of production." Since irrigation increases yields two to fourfold, irrigating one acre of land "is the same as 'finding' 1–3 additional [acres] of cropland even when the irrigated area had been cultivated before."

Both Hopper and Johnson, as well as many other authorities (for example, Bauer 1981, pp. 177–84; Berry 1977, p. 271; and Krauss 1983, p. 160) agree that the principal cause of low yields and the slow rate of improvement in Third World agriculture stems from the systematic exploitation of Third World farmers by their own governments. For example, Johnson (p. 99) notes that in Egypt

> the farm price of rice was 30 percent of the world price and for cotton, about 40 percent; in Yemen, Brazil, Malawi, and Upper Volta the farm price of cotton was just 70 percent of the world price; in India, the Philippines, Bangladesh, Pakistan, and Senegal rice prices were 30 percent below the world price, but Tanzanian farmers would have been overjoyed with such treatment since their rice prices were 70 percent *below* the world price.

This implicit tax on Third World farmers clearly discourages agricultural output.

Contrary to Paul Ehrlich's prediction of increasing poverty and world hunger, the UNFAO data clearly show a continuing, long-term improvement in the food situation of even the less-developed countries. Moreover, the large amount of uncultivated arable land in the less-developed countries raises the possibility of dramatic acceleration in the rate of this improvement, especially if equitable farm policies were adopted.

Second, the available data show that for at least the last 100 years the prices of practically all minerals have declined, despite short-term fluctuations. This is true whether one compares mineral prices to wages for labor (see Table 2) or uses the consumer price index (Simon 1981b, pp. 96–97, 349–51). The most recent studies (Barnett et al. 1984, pp. 316–38) show that there is no basis for the popular belief that these long-term trends were reversed during the 1970s. In contrast to the predictions of Ehrlich and other Neomalthusians, the logical inference is that resources are actually becoming less scarce over time.

Third, instead of a decade of "environmental deterioration," the 1970s were actually years of environmental improvement, especially in the United States. According to data supplied by the Environmental Protection Agency (U.S. Dept. of Commerce 1983, p. 211), 112.8 million metric tons of carbon monoxide were emitted into the atmo-

TABLE 2

PRICES OF IMPORTANT MINERALS RELATIVE TO LABOR COSTS

Mineral	1900	1920	1940	1950	1960	1970
Coal	459	451	189	208	111	100
Copper	785	226	121	99	82	100
Iron	620	287	144	112	120	100
Phosphorus	—	—	—	130	120	100
Molybdenum	—	—	—	142	108	100
Lead	788	388	204	228	114	100
Zinc	794	400	272	256	126	100
Sulphur	—	—	—	215	145	100
Aluminum	3,150	859	287	166	134	100
Gold	—	—	595	258	143	100
Crude Petroleum	1,034	726	198	213	135	100

SOURCE: Nordhaus (1974, p. 24).

sphere in the United States in 1970. In 1981 that figure was 90.5, a decline of nearly 20 percent. Sulfur oxides fell from 26.4 million metric tons in 1970 to 22.5 in 1981. Volatile organic compounds fell from 27.2 million metric tons in 1970 to 21.3 in 1981. And particulates—suspended particles of smoke, dust, fumes, and droplets of viscous liquid—registered the largest improvement, going from 17.9 million metric tons in 1970 to only 8.5 million in 1981, a decline of over 50 percent. Nitrogen oxides were the only pollutant to increase, going from 17.6 million metric tons in 1970 to 19.5 in 1981, a modest increase at worst. The results of this reduction in emissions are evident in the EPA's Pollutant Standards Index (PSI), which is designed to measure the overall quality of air in metropolitan areas. In 1979 there were 24 readings in which pollution was at a "hazardous" level in one or more metropolitan areas. The PSI recorded 364 readings in the "very unhealthful" range. By 1981 the number of "hazardous" recordings fell to 3. The number of "very unhealthful" readings declined to 262 (U.S. Dept. of Commerce 1983, p. 211). The National Commission on Air Quality (p. 1) reported in 1981 that during the past decade "the absolute level of improvement for most widespread pollutants has been significant." The conclusions of the U.S. Council on Environmental Quality (CEQ) were similar (1981, p. 271): "Air quality has been improving. The number of air pollution alerts has been reduced. The large amounts of chemical pollutants coming from industry have been reduced, and emissions from the worst offenders have been markedly curtailed."

Water pollution also showed marked improvement throughout the 1970s. According to the CEQ (Simon 1981a, p. 9), the number of monitoring stations reporting "good" drinking water rose from 42 percent in 1961 to 61 percent in 1974; those reporting "fair" drinking water rose from 28 percent to 34 percent. On the other hand, the number of stations reporting either "very poor" or "severe" water dropped from 9 to 1 percent. This progress appears to have continued during the latter half of the decade. A rough indication of water quality can be obtained by examining the "violation rate" for major water pollutants. This rate "represents the proportion of all measurements of a specific water quality pollutant which exceeds the 'violation level' for that pollutant." According to the CEQ (U.S. Dept. of Commerce 1983, p. 209) the violation rate for fecal chloroform bacteria was 46 percent in 1973 and 30 percent in 1981. It fell from 6 percent for dissolved oxygen in 1973 to 4 percent in 1981, while phosphorus fell from 57 to 49 percent. The violation rate for lead dropped from 13 percent in 1979 (the first year for which data are available) to only 3 percent in 1981, and cadmium fell from 4 to 1 percent between 1979 and 1981. Overall, the 1970s were years of air and water improvement. One can argue that the method adopted to control pollution is costly and inefficient, but that, of course, is a separate issue. The Neomalthusians' prediction that the 1970s would be a decade of environmental deterioration is contrary to the data.

Fourth, instead of declining during the 1970s, life expectancy has continued its steady increase and currently stands at 74.5 years for the United States and 59.2 years for the world as a whole (U.S. Dept. of Commerce 1983, pp. 73 and 856), compared to 69.0 and 47.2 years, respectively, in 1950–55.

Fifth, not only are the oceans not dead, the world fish catch increased from 146.6 billion pounds in 1975 to 159.2 billion in 1980 (U.S. Dept. of Commerce 1983, p. 703).

Finally, there is Ehrlich's prediction of a Midwestern desert. Table 3 shows rather clearly that, first, the yield per acre in the Midwestern states for the corn, wheat, and soybean crops has consistently been above the national average, and, second, yield per acre was higher for all three crops in 1980 than it was in 1970. These patterns show no sign of changing. Unless increasing yield per acre is the harbinger of new deserts, Ehrlich's prediction is flatly refuted by the data.[2]

---

[2] Many experts believe that outputs will increase even faster in the future, especially in the industrialized countries. One reason is biotechnology, the science of breeding plants and animals. Among the results expected before the turn of the century are crops that need no fertilizer, plants that resist disease, drought, and insects, leaner and more nutritious grain, and pigs and cattle that grow twice as fast (Ritter, 1984, sec, 6, pp.1–2).

TABLE 3

PRINCIPAL MIDWESTERN CROPS: YIELDS PER ACRE

	1961–65	1966–70	1970	1975	1980	1982
			Corn			
U.S. average	66.4	78.1	72.4	86.2	91.0	115.0
Illinois	NA	NA	74.0	116.0	93.0	134.0
Indiana	NA	NA	74.0	98.0	96.0	129.0
Ohio	NA	NA	77.0	92.0	110.0	117.0
			Wheat			
U.S. average	25.3	28.3	31.0	30.7	33.4	35.6
Illinois	NA	NA	36.0	39.0	48.0	45.0
Indiana	NA	NA	38.5	43.0	49.0	43.0
Ohio	NA	NA	37.0	42.0	49.0	44.0
			Soybeans			
U.S. average	24.2	26.2	26.7	28.8	26.4	32.2
Illinois	NA	NA	NA	36.0	33.5	39.0
Indiana	NA	NA	NA	33.5	36.0	40.0
Ohio	NA	NA	NA	33.0	36.0	37.0

SOURCES: U.S. Dept. of Commerce (1972, pp. 608–11; 1977, pp. 700–708; 1983, pp. 676–80).

Special mention should be made of oil since the price rise in the early 1970s spawned a barrage of stories about the coming exhaustion of this vital nonrenewable resource. This conclusion is surely erroneous. For while the *price* of Mideast oil rose to about $35.00 per barrel, the *cost* of extraction remained at between $0.05 and $0.15 per barrel. The higher prices have nothing to do with increasing extraction costs due to the exhaustion of oil reserves. They are the result of the confluence of political factors such as, internationally, the OPEC cartel, and domestically, price controls on natural gas and other energy products which, by reducing domestic supplies, inadvertently played into the hands of OPEC (Osterfeld 1984, pp. 366–70; Osborne 1984, pp. 86–102).

Mistakes in the Neomalthusian Methodology

The key question is: How have the Neomalthusians gone so far wrong? Most of their doomsday calculations are based on the supposition that known reserves have calculable exhaustion dates. Thus we are told, for example, that at current consumption rates the known reserves of copper will be exhausted in 2010, petroleum in 1995, and tin in 2003. Such a supposition is seriously flawed, and its guarantee of predictions of imminent exhaustion for whatever resource one wishes to examine cannot be sustained.

Known reserves are only a very small part of the available resource base. According to Ridker and Cecelski (1982, p. 595), they are similar to inventories, "and since exploration and development are costly, little effort is made to find proof of new resources if what is already known is considered adequate to meet demands for the next ten to twenty years." The World Bank made the same point (Beckerman 1975, p. 175): "The reason we do not know the absolute limits of the resources we have is quite simple. . . . we do not know because no one has yet found it necessary to know and therefore went about taking an accurate inventory." In short, we find resources as we need them. Hence, calculations based on current reserves produce "exhaustion dates" that are unrealistically close-at-hand.

A flaw in the notion of exhaustion dates is that it ignores the "feedback mechanism." As current reserves are consumed and the exhaustion date for a given resource draws near, the price will begin to rise. This will reduce consumption and encourage recycling, exploration for new reserves, and the development of technological improvements that will make the mining of lower-grade ores economically feasible. So despite the predictions of imminent depletion of numerous resources over the years, their exhaustion dates never get any closer. In fact, since improved methods of mineral detection have increased the known reserves of nearly all raw materials, exhaustion dates have actually receded (see Table 4).

Put differently, there are always "shortages" of raw materials, just as there are always "shortages" of shoes, toilet paper, and automobiles. These shortages are independent of any particular economic system. Economic institutions, says Thomas Sowell (1980, p. 45), "exist to introduce elements of rationality, or efficiency, into the use of inputs and outputs." A rational policy is one that insures that adjustments to changing economic data are made incrementally, that is, "at the margin." As the supply of a resource increases relative to its demand, the exhaustion date will recede. It is rational for factors of production to be diverted from areas where the exhaustion date is relatively remote, and thus the demand for additional units of the resource is relatively less urgent, into areas where the exhaustion date is closer at hand and thus demand for additional units of those resources (or finished goods or services) is relatively more urgent (Stroup and Baden 1983, pp. 4–5). Consider, says Sowell (1980, p. 46), the consequences of alternative decision-making processes.

> The alternative would be to completely satisfy all of some category of needs—the most urgent, the moderately important, and the

TABLE 4

CHANGES IN "KNOWN RESERVES" 1950–1970

Ore	*Known Reserves in 1950 (1,000 Metric Tons)*	*Known Reserves in 1970 (1,000 Metric Tons)*	*Percentage Increase*
Iron	19,000,000	251,000,000	1,321
Manganese	500,000	635,000	27
Chromite	100,000	775,000	675
Tungsten	1,903	1,328	–30
Copper	100,000	279,000	179
Lead	40,000	86,000	115
Zinc	70,000	113,000	61
Tin	6,000	6,600	10
Bauxite	1,400,000	5,300,000	279
Potash	5,000,000	118,000,000	2,360
Phosphates	26,000,000	1,178,000,000	4,430
Oil	75,000,000	455,000,000	507

SOURCE: Kahn et al. (1976, p. 92).

> trivially marginal—thereby leaving still more unsatisfied (and more urgent) needs unmet elsewhere in the economy. . . . The mundane fact of insufficiency must be insisted upon and reiterated because so many discussions of "unmet needs" proceed as if "better" policies, practices, or attitudes would "solve" the problem at hand without creating deficiencies elsewhere.

The Neomalthusians tend to ignore the marginalist principle and take an all-or-nothing view of the world. It is precisely this oversight that has led them to derive such erroneously alarmist conclusions from such mundane concepts as "known reserves" and "exhaustion dates."

Growing Abundance of Resources

The prospect of "running out" of a resource such as oil is not a matter of decades, as the doomsdayers would have us believe, but lies hundreds, perhaps thousands of years ahead for *all* vital resources. The World Bank has estimated the stock of metals in the top mile of the earth's crust to be a million times as great as the present known reserves. Beckerman (1975, p. 196) has calculated that "this means that we have enough to last about one hundred million years." Kahn et al. (1976, p. 102) determined that "over 95 percent of the world demand is for five metals (iron, aluminum, silicon, magnesium, and titanium) which are not considered exhaustible." Another 4.85 percent of the world demand is for seven metals—copper, zinc, manganese, chromium, lead, nickel, and tin—which are "probably inexhaustible." Thus, 99.9 percent of the world demand is for metals whose supply is either "clearly" or "probably inexhaustible."

Goeller and Weinberg (1976, pp. 683–89) have shown that of the 13 most widely used resources, only fossil fuels and phosphorus are "not essentially inexhaustible." Phosphorus, they write, has a "present resource-to-demand ratio [of] 500 years for world reserves and an additional 800 years for potential resources." They further contend that the development of alternatives to fossil fuels, such as the fission breeder, fusion, or solar energy, would provide the world with "inexhaustible sources of energy." Osborne (1984, pp. 86–87) has noted that although many experts were predicting the end of methane supplies by the late 1980s, the "drilling binge" that followed partial decontrol of natural gas prices in 1978 produced a "glut of gas so severe" that, if these sources could be extracted at commercially feasible prices, there would be enough "to supply all apparent needs for a thousand years."

Nordhaus (1974, pp. 22–26) arrived at similar conclusions. Using U.S. Geological Survey data, he presented resource availability according to three measures: "known reserves," "ultimate recoverable resources," and "crusted abundance" (the total amount estimated in the earth's crust). Dividing each measure by the current consumption rates, Nordhaus provided an estimate of resource availability in years (Table 5). He dismissed the "known reserve" measure as "unduly pessimistic" for reasons discussed above. The "crustal abundance" measure was rejected as "unduly optimistic because it assumes that everything [in the earth's crust] can be recovered." "Ultimate recoverable resources" therefore was left as the most reasonable measure. According to Nordhaus, the Neomalthusian concern about imminent resource exhaustion is without foundation, especially since these findings are based on the state of existing technology and "do not take into account the economic feasibility of mining lower grade ores as prices rise or techniques improve." The clear evidence, Nordhaus concludes, "is that the future will *not* be limited by the sheer availability of important materials." Nordhaus (1974, p. 24) reached similar conclusions for energy. At the ratio of reserves to 1970 consumption rates, he estimates there would be enough fossil fuel to last for over 500 years (Table 6, line 1). And "even with only current technology there are resources for more than 8,000 years at the current rate of consumption [see Table 6, line 2]. With breeder reactors, and more dramatically with a fusion technology, there is virtually unlimited energy available."

TABLE 5

MEASURES OF MINERAL CONSUMPTION (IN YEARS)

Mineral	*Known Reserves ÷ Annual Consumption*	*U.S. Geological Survey's Estimates of "Ultimate Recoverable Resources" (= 1% of Materials in Top Kilometer of Earth's Crust) ÷ Annual Consumption*	*Amount Estimated in Earth's Crust ÷ Annual Consumption*
Copper	45	340	242,000,000
Iron	117	2,657	1,815,000,000
Phosphorus	481	1,601	870,000,000
Molybdenum	65	630	422,000,000
Lead	10	162	85,000,000
Zinc	21	618	409,000,000
Sulphur	30	6,897	NA
Uranium	50	8,455	1,855,000,000
Aluminum	23	68,066	38,500,000,000
Gold	9	102	57,000,000

SOURCE: Nordhaus (1974, p. 23).

TABLE 6

ENERGY RESOURCE-CONSUMPTION RATIO, 1970

Energy/Technology	Years
Fossil fuels only	520
Fossil fuels plus current nuclear technology	8,400
Fossil fuels, current nuclear and breeder technology	1.1×10^6
Fossil fuels, current nuclear, breeder, and fusion technology	5.3×10^{10}

SOURCE: Nordhaus (1974, p. 24).

Finally, even the *Global 2000 Report* (USCEQ et al. 1982, p. 2) acknowledges that the depletion of important resources is not an immediate concern: "[T]he world's finite fuel resources—coal, oil, gas, oil shale, tar sands, and uranium—are theoretically sufficient for centuries. . . . Nonfuel mineral resources generally appear sufficient to meet projected demands through 2000." Moreover, geothermal energy, that is, energy obtained by tapping into the heat generated by radioactive decay in the earth's interior, represents a virtually unlimited alternative source. Advances in drilling technology are expected to make this a reality by the mid-21st century (Beckermann 1984, p. 425).

One of the pillars of Neomalthusian thought is the second law of thermodynamics, the law of entropy (Georgescu-Roegen 1971, 1976; Daly 1974, 1979; Commoner 1976). This law points out that although the quantity of energy in the universe is constant, the "usable stuff" is low-entropy energy. As energy is used, it is transformed from useful (low-entropy) to useless (high-entropy) energy. This process is irreversible. The law of entropy, say the Neomalthusians, proves that the world is headed, inexorably, toward the catastrophe of energy depletion. But if the studies of Nordhaus, Kahn, and Goeller and Weinberg are correct, we have enough fossil fuels in the earth to last for over five centuries. With current nuclear and breeder technology there is enough energy to last for about a million years. And then, of course, there is the sun. While the Neomalthusians admit that the sun will be a steady source of low entropy for five to six billion years, they emphasize that solar energy "is strictly limited in the rate at which it reaches the earth" (Daly 1979, p. 74). But the amount of direct solar energy absorbed by the earth is about 4,500 times the world's current energy consumption rate (Beckermann 1984, p. 415). The point is not that the law of entropy is incorrect but that the conditions which would make the law a concern in the area of energy accessibility are so remote as to render the law irrelevant for the foreseeable future. The real issue concerning direct solar energy lies not with its volume but with its "diluteness," or density. Beckermann (1984, p. 418) notes that the history of man's energy use "is a history of increasing energy concentrations." The problem is that direct solar energy is so dilute—about one kilowatt of energy per square meter under ideal circumstances—that one must question whether direct solar energy can ever become an economically viable energy source. Even if this should prove the case, the difficulty would lay with the diluteness of solar energy and not, as the Neomalthusians would have it, with "the rate at which it reaches the earth."

Science, Technology, and Resources

If the depletion of the earth's stock of a vital raw material such as oil ever did appear imminent, we still can be confident that it would not have serious consequences, provided the market is unhampered. As the exhaustion date drew near and exploration failed to yield new reserves, the price of oil would rise. Not only would this stretch the supply by encouraging conservation, but the rising price would also encourage the search for both resource substitutes, such as nuclear, solar, and geothermal energy, and factor substitutes, such as capital and labor. It would also stimulate the search for technological developments that might permit us not just to produce oil synthetically but to produce it at economically feasible prices. In such a case, the exhaustion of even a key natural resource would be, as Solow (1974, p. 11) has phrased it, "just an event, not a catastrophe."[3]

Kahn et al. (1976, p. 59) observed that the steep increase in the price of oil in the early 1970s generated just such responses. The oil crisis "represented an energy watershed," they wrote, "but it was not a watershed from abundance to scarcity, or even from cheap to expensive, but rather from cheap to inexpensive." They argued that the effect of the sudden price hike "was to increase the rate at which new energy came on the market and to decrease the rate at which energy was used—that is, the cartel's moves actually decreased the possibility of future energy shortages." This conclusion was verified by Louw (1985, pp. 14–18). The proven reserves of both coal and gas are at all time highs, with additional finds being made on a regular basis. But "alternative" energy sources are being developed so rapidly that, says Louw, "far from the world running out of coal or oil, it is a very real prospect that alternate sources will render them redundant. Oil may again become an 'irritable, nasty, black ooze' and coal a 'dirty black rock.' " And finally, there is no reason to rule out future exploration of the moon or other planets to replenish the earth's energy supplies.

---

[3] A staple of the Neomalthusian position is that we are consuming resources "too quickly." The logic is that since life is short, individuals discount future for present consumption. But "society" does not die. Therefore, in social decision making "there is no excuse for treating generations unequally." Hence, "we ought to act as if the social rate of time preference were zero." Solow (1974, p. 9) says that "I find that reasoning persuasive." This is a curious conclusion for he also states that "if the future is anything like the past," resources will become more abundant over time, they will be used with increasing efficiency and, with technological improvements, substitutions will become progressively easier so that the exhaustion of a natural resource will be "just an event," and an unlikely one at that. If so, why should we be concerned about saving resources for future generations? Indeed to the extent that "intergenerational equity" is a concern at all it would be that present generations are being exploited by future generations, that is, that we are consuming resources "too slowly."

The earth is a physically finite place, but its resources are neither fixed nor finite. The concept of resources is a dynamic one. While it is common knowledge that technology consumes resources, a more important fact is often overlooked: Technological advances permit us to use existing resources more efficiently. For example, in 1900 the lowest grade of copper ore economically mineable was about 3 percent. Today, the cutoff point has fallen to 0.35 percent. Similarly, although much of the coal closest to the surface has been extracted, advances in mining technology have actually reduced the cost of obtaining coal, despite having to go much deeper to extract it. More important, technological advances actually create resources by finding uses for previously useless materials. Uranium is one example; hydroelectric power is another. But oil is the most dramatic. Prior to the mid-19th century oil was a liability, and land known to possess this slimy ooze was worth very little. Only with the dawn of the machine age did oil become a resource. Aluminum is another example. Throughout the 19th century aluminum was a precious metal on a par with gold or silver, but with advances in technology, aluminum can now be extracted from bauxite very cheaply. Bauxite and other materials containing aluminum deposits, such as clay and shale, are so plentiful that we now have nearly inexhaustible supplies of aluminum at relatively low prices (Beckerman 1975, pp. 185–86).

By enabling us to find and use resources more efficiently and by discovering uses for heretofore useless materials, technology creates resources. Cook (1976, pp. 677–78) put this idea quite succinctly:

> During the past 150 years large increases in the earth-resource base of industrialized society have been attained. By increasing the efficiencies of discovery, recovery, processing, and application of such resources, we have been able to find and exploit leaner, deeper,

> and more remote deposits. By discovering and developing new methods of utilizing previously worthless materials we have created resources where none existed.

Although the notion that resources are becoming more abundant may appear counter to common sense, it has solid empirical foundations. How else can one explain that by whatever index one chooses, the costs of all important raw materials have been declining for at least the last 100 years? But while technology creates resources, it must be remembered that technology itself is created by the human mind, which Julian Simon (1981b) aptly dubs "the ultimate resource." The supply of resources, therefore, is limited only by the ingenuity of the human mind. Although this too runs counter to conventional wisdom, there is strong empirical support for it. That the possession of physical resources is not necessary for economic development is illustrated by Hong Kong, Singapore, and Japan, all of which have few physical resources yet highly developed economies. That the possession of physical resources does not guarantee economic development is demonstrated by numerous African countries that possess an abundance of physical resources yet are economically backward.

The flaw in the Neomalthusian outlook is the seemingly plausible but erroneous assumption that the stock of resources is fixed, that the size of the resource pie is determined by nature. What is important, however, is not the actual physical substance itself, but the services we are able to derive from it. As Simon (1981b, pp. 46–47) notes:

> What is relevant to us is not whether we can find any lead in existing lead mines but whether we can have the services of lead batteries at a reasonable price; it does not matter to us whether this is accomplished by recycling lead, by making batteries that last forever, or by replacing lead batteries with another contraption.

Viewing resources as services rather than stocks leads to the important conclusion that the resource pie is not fixed but changeable, and thus there is no inherent reason for the stock of resources to diminish over time. On the contrary, since resources are a function of knowledge and since our "stock of knowledge" has increased over time, the logical inference is for resources to become increasingly abundant. As we have seen, the data clearly show this to be the case.

Resources, the Market, and the Entrepreneur

Scientists and technicians, of course, are not merely sitting around waiting for "shortages" to appear. How is it that they are directed to the areas where their services are most needed? Here we come to a very different but perhaps even more indispensable type of information—that provided through the market—and to the vital but often neglected role of the entrepreneur acting on that information.

In contrast to scientific and technological knowledge, which may be viewed as an expanding body or corpus, market information is fleeting and ever changing. This is so by its very nature. Consumer wants change. It would be of little interest or help to a businessman concerned with consumer wants in the 1980s to be told that crew cuts, string ties, and Bill Haley were popular in the 1950s. Other economic data underlying supply and demand schedules and costs and prices are just as ephemeral. In contrast to a particular scientific field such as physics or even economics, where one can be fairly confident that the general body of principles will not change to any significant degree, economic data must be forever learned anew. But these changes, as haphazard and chaotic as they may appear on the surface, are connected, or coordinated, spontaneously as it were, through the market. The entire process is determined by the consumer.

Profits and losses are indicators of maladjustments; the structure of production is not completely synchronized with consumers' preferences. Losses occur because what is being produced is not what consumers intensely desire at that particular time and price. This means that, in the eyes of the consumers, those factors can be more usefully employed elsewhere. The adjustment will be made because production cannot be continued at a loss. Conversely, profits are an indication that, in the eyes of consumers, too little of a good is being produced. Profits also serve to attract resources into the field. Thus, it is the changes in relative prices, caused by consumer buying and abstention from buying, that determine the entire process, from the search for raw materials to the completion of the finished product. Changes in consumers' preferences cause changes in profit margins and thus in the allocation of resources. In a free market, therefore, profits automatically are correlated with serving consumers while losses indicate a failure to serve.

We have discussed the process by which the economy adjusts to consumer demand. But why does this process occur? If the consumer is the "sovereign" in the system, the entrepreneur is the motor. It is the misallocation of resources resulting from changes in economic data that creates profit opportunities. It is the entrepreneur who, by spotting and exploiting these opportunities, moves the economy toward a more efficient allocation of resources. As Kirzner (1973, pp. 73–75, 81) puts it:

> The important feature of entrepreneurship is not so much the ability to break away from routine as the ability to perceive new opportunities which others have not yet noticed. Entrepreneurship . . . is

> not so much the introduction of new products or of new techniques of production as the ability to *see* where new products have become unsuspectedly valuable to consumers and where new methods of production have, unknown to others, become feasible.

"It is only through the entrepreneur, that changes can arise," he says. The entrepreneur "brings into *mutual adjustment* those discordant elements which resulted from prior ignorance."

The market and the entrepreneur are the opposite sides of the same coin. Neither can function without the other. Their critical role in the area of resources is evident. Scientists and technicians are drawn into those areas where they are most needed because entrepreneurs, perceiving the signals conveyed by price changes in the market, bid those scientists and technicians away from areas where their contributions are less urgently needed and into areas where they are more intensely needed.

It is instructive to consider the only alternative to the market: the centrally planned, nonmarket economy. Without the market, there would be neither signals conveyed by market prices nor profit-seeking entrepreneurs to act on those signals. The closest example is the USSR, where all decisions, at least in theory, are made by a single central planning board, Gosplan. The Soviet economy is notoriously inefficient and would be even more so if it were not for extensive black markets. The standard of living is the lowest of any developed country, its products are notoriously shoddy, and the overall growth rate has declined in every period since 1950 (Goldman, 1983, pp. 46–47).

Since consumer goods are directly valued, their demand can be at least roughly determined by the planners. But resources are valued only indirectly, that is, for their contribution to the production of a final good. This renders the problem of the efficient allocation of resources particularly acute. The number of alternative technologies and different ways of utilizing scarce resources for an economy the size of the Soviet Union is, literally, in the millions. Without valid price data, the choices that Gosplan makes are little more than guesses. There is little doubt that this is the case. Producers in the Soviet Union have no direct connection to consumers. They are assigned quantity production targets of so much steel, so many nails, and so on, and are then evaluated according to whether they have met their targets. Because these targets rarely coincide with "improvement in the country's economic well-being," the planning process, notes Goldman, results in a great deal of waste of both human labor and natural resources. A few examples will make this clear.

When production targets are given in weights, factory managers in the nail industry produce large construction nails and no small finishing nails. Shoe manufacturers produce adult shoes but no children's shoes. When production targets are switched from weights to quantities the result is a glut of finishing nails and children's shoes and a shortage of construction nails and adult shoes. The same sort of distortions and waste occur regularly in the area of energy resources. The Soviet oil industry, for example, is evaluated not according to the amount of oil discovered but rather on the number of meters drilled each month, making the discovery of oil only incidental to the primary goal of drilling the number of meters assigned as the quota. Since the deeper the well the slower the drilling process, geologists have responded by digging numerous wells too shallow to hit oil. As *Pravda* (Goldman 1983, pp. 38–39) has noted, "There are geological expeditions that have not discovered a valuable deposit for many years but are counted among the successful expeditions because they have fulfilled their assignment in terms of meters." The result has been the waste of resources and labor.

Another incentive method used by the Soviet planners is the *valovaia produktsia,* or "val," which is intended to be a measure of "gross ruble-output." Since plant managers are rewarded for increasing the gross value of their factories' outputs, they respond by intentionally increasing production costs. Hence, the val has provided an incentive to squander scarce resources by rewarding managers for using large quantities of the most expensive resources they can find.

These problems impede the search for substitutes. As Goldman (1983, p. 42) notes,

> Soviet managers seem . . . to have an almost phobic reaction to innovation. The main reason is that the Soviet manager usually has little to gain and much to lose from innovation. In order to revamp his production line to make way for a new product, the manager usually has to halt production. This means that he would be unable to fulfill his present production targets, and thus he will not qualify for his bonus. At the same time, the Soviet incentive system does not provide any substantial increase in bonuses for the production of new goods. Thus there is little in the way of extra inducement for innovation, but considerable disincentive.

In brief, central planning as it has worked in the USSR is notoriously inefficient. This inefficiency is not a result of incompetence. For even if the factor allocation established in the plan were optimal, changes in data during the plan period would mean that that allocation would no longer be optimal at the end of the period. Thus, central planning is inherently inefficient. This inefficiency is particularly acute in the area of energy resources where the incentives created by the plan encourage the squandering of these resources (such as oil) while simultaneously impeding innovations and the search for substitutes.

Government and Markets: Poverty and Prosperity

The foregoing raises a significant issue. If prosperity and economic development are neither negatively correlated with population growth or density nor positively correlated with arable land or physical resource endowment, how does one account for the existence of a handful of prosperous developed nations and a large number of poor undeveloped countries? While the causes of economic development are complex, there is one unmistakable correlation: those countries with freer markets have a higher per capita income and a faster rate of growth than those countries where markets are more constrained.

Table 7 contrasts the performance of several market-oriented countries with the more controlled economies of their sister states. The results are dramatic and can hardly be attributed to accident or coincidence.

The same relation holds true when one looks at growth rates. As Table 8 shows, the relatively low tax, private sector countries have clearly outperformed the relatively high tax, public sector economies.

The superiority of the free market over the controlled economy is also demonstrated by an examination of nutritional data. "Of all the 'socialist experiments,' " says Eberstadt (1981, p. 46), "only two provide evidence of nutritional results noticeably better than their non-Communist neighbors: Soviet Asia and Cuba." Yet the former's neighbors, such as Afghanistan, can hardly be termed market oriented. And Soviet aid to Cuba is approximately 50 percent of Cuba's internally generated GNP. "It could not possibly survive, let alone function," says Eberstadt, "without enormous annual subsidies from

TABLE 7

PER CAPITA INCOME FOR 1979
(IN U.S. DOLLARS)

Market Economies		Controlled Economies	
Taiwan	1,667	China	576
South Korea	1,385	North Korea	750
West Germany	10,670	East Germany	6,318
Ivory Coast	916	Ghana (Gold Coast)	411

SOURCE: *Economic Handbook of the World 1981.*

TABLE 8

AVERAGE ANNUAL GROWTH OF GROSS DOMESTIC PRODUCT

	Private Sector Economies			Public Sector Economies	
	1960–70	1970–79		1960–70	1970–79
Hong Kong	10.0	9.4	India	3.4	3.4
Singapore	8.8	8.4	Uruguay	1.2	2.5
Thailand	8.2	7.7	Ghana	2.1	−0.1
South Korea	8.6	10.3	Chad	0.5	−0.2
Ivory Coast	8.0	6.7	Zaire	3.6	−0.7

SOURCE: World Bank (1981, pp. 126–37).

the Soviet Union." In brief, the correlation of poverty and hunger with government planning and regulation, and of economic development and nutritional progress with economic freedom, is quite remarkable.

Institutional Incentives and Economic Stagnation

Throughout much of the Third World one finds high minimum wages, state-created monopolies, rigid licensing restrictions, tariffs in excess of 100 percent, widespread nationalizations, and the systematic exploitation of farmers. The policy implications are clear: These interventions have created major distortions resulting in a serious misallocation of resources; they have frightened off investors, domestic as well as foreign; and they have impeded development, which in turn has exacerbated, in fact *caused,* the population pressures experienced by many Third World countries.

Despite their economic irrationality, there is little hope that many of these programs will be dismantled. The persistence of the highly interventionist policies of so many Third World governments, especially in the face of their continuing failure to generate sustained growth, suggests that there is more involved than misguided idealism. These policies almost invariably benefit small but politically powerful urban elites. It is naive to expect these elites voluntarily to dismantle these programs, thereby surrendering the substantial economic and political benefits accruing to their positions. It is here that we come to the important but much neglected part that political corruption plays in retarding Third World development.

Osusu, the Ashanti word for "corruption," pervades the Third World. For example, President Mobutu Sese Seko of Zaire, whose personal fortune is estimated at $4 billion, is widely suspected of embezzling

$1 billion from the state treasury. The former head of the Central African Republic, Jean-Bedel Bokassa, had himself declared president-for-life in 1976 and then staged a $20 million ceremony in honor of the event. Bokassa was deposed two years later. President El Hadj Oman Bongo of Gabon squandered $27 million on the construction of a conference center to be used solely for the 1977 meeting of the Organization of African Unity. But corruption is not confined to heads of state. "Even in the poorest African capitals," *Time* ("A Continent . . ." 1984, p. 28) reports,

> government officials can be seen in convoys of Mercedes-Benz limousines scattering cyclists and pedestrians as they pass. Owning a Mercedes is so potent an African status symbol that in East Africa a Swahili word was coined to describe the elite that drives them: *wabenzi,* literally, men of the Mercedes-Benz.

Corruption, Senegal's President Abdou Diouf lamented, "is endemic to the continent." But it is by no means restricted to Africa. Nearly all of the $3 billion given to Bangladesh in the early 1970s was pocketed by Mujib and his clan (Mehta 1978, p. 134). And the Wisers (1971, pp. 100–16) refer to the pervasive "opportunities for corruption" that accrue to government appointed village headmen in India. "If you were to take one of the most harmless men in your village," a young Indian confided to them, "and put him in the watchman's place, he would be a rascal within six months."

Of course corruption is also found outside the Third World. But what is noteworthy about Third World corruption is its sheer magnitude and pervasiveness. The causes are complex but certainly a major reason is that, in contrast to the free press, open dissent, and competitive elections found in much of the West, Third World countries are characterized by censorship, political oppression, and one-party regimes or military dictatorships. For example, only 7 of the 41 major black African nations permit political opposition. Seventeen are military regimes, and another 17 are single-party states.

What makes corruption so extensive in the Third World is that, in the absence of effective checks on the power of the rulers, they are free to use government largess as they wish. Few can resist such temptation. But since the benefits seized by the elites must be paid for by others,those benefits are in the nature of a tax, which, like any tax, discourages output and retards development. It may seem therefore that, short of physical intervention, there is little or nothing the Western nations can do to induce sensible reforms.

In fact, however, foreign aid, or "Official Development Assistance" (ODA), to the poorest of the Third World countries increased dramatically during the 1970s, rising from 4.1 percent of their combined gross domestic product in 1971 to 10.3 percent in 1978. The role of ODA in perpetuating Third World poverty by facilitating and encouraging corruption by Third World leaders cannot be overlooked. Nor can its potential as a vehicle for inducing reforms without intervention be ignored.

ODA can and has been criticized on strictly economic grounds. It politicizes life in the recipient country, thereby diverting energy from economic to political activities (Bauer and Yamey 1983); it blurs lines of investment and distorts cost data, resulting in the waste of resources (Osterfeld 1982); it reduces pressure on the recipient government to maintain an environment favorable to private enterprise, thereby discouraging private investment (Friedman 1958, p. 207); and it retards the development of those attitudes—thrift, industry, and self-reliance—that are essential for economic health (Bauer 1972, pp. 97–98).

More important, however, is the incentive structure created by current aid programs. Foreign aid "goes not to the pitiable figures we see on aid posters or in aid advertisements," Bauer and Yamey point out (1983, p. 125), "it goes to their rulers." ODA is dispensed on the basis of need, usually defined in terms of a country's per capita income. A prime example is the International Development Association of the World Bank, which restricts aid to those countries with a per capita income of less than $730. The "loans" are made to governments at highly concessional terms: pay-back periods of 50 to 100 years and zero percent interest, although a 0.75 percent service charge is assessed. The result is clear: ODA provides Third World rulers with an incentive to perpetuate the poverty of their subjects.

There is little doubt that this is the case. Throughout the Third World one finds entire occupations being outlawed and hardworking and industrious groups subjected to brutal treatment ranging from discrimination to exclusion from choice occupations to outright slaughter. An example of the former is Mobutu's expulsion of traders or middlemen that promptly reduced Zaire's per capita income, thereby qualifying Zaire for increased "assistance" (Bauer and Yamey 1983, p. 125). Examples of the latter include the brutal mistreatment of economically wealthy but politically weak minorities in Algeria, Burma, Burundi, Egypt, Ethiopia, Ghana, Indonesia, Iraq, Kenya, Malaysia, Nigeria, Sri Lanka, Tanzania, Uganda, Zaire, and Zambia. "Because the victims' incomes were above the national average," says Krauss (1983, p. 158), "their maltreatment promptly reduced average incomes and thereby widened income differences between these countries and the West." The result, he points out, is that the self-inflicted economic deterioration qualified these countries for additional "aid."[4]

Since "bureaucratic success" is measured by the size of the agency's budget, or, in the case of transferring organizations, by the volume of loans dispensed, these agencies have far more incentive to increase the amount of wealth transferred than to be concerned about how it is used (Sowell 1983, p. 238). Both the World Bank and the IDA are examples. The Bank's authorized capital increased from an initial $12 billion in 1944 to $86.4 billion in 1981. The IDA's increased from an initial $916 million in 1960 to $12 billion in 1981.

Some observers, however, have defended World Bank and IDA activities. Ayres (1983, pp. 15, 37 and 63), for example, argues that while there have been difficulties, World Bank and IDA loans have benefited the world's poor, and any curtailment would cause "societies in transition" to suffer. Yet one finds little in the way of concrete evidence to support his conclusion. On the contrary, there are numerous references to such things as "benefit deflections" (pp. 103, 124 and 193) and "shortfalls" (p. 126). Ayres states that the World Bank always "seeks assurances from the recipient country" about the way the loan will be used, but then observes that "the Bank can obtain all of the assurances it wants, but it is up to the recipient country to make good on them, and the Bank does not always possess the leverage or supervisory capability for seeing to this" (pp. 43–44).

Elsewhere he says that "the political elite in most recipient countries does not care about the poor majority. Where there is the absence of political will and commitment it is difficult for the Bank to be effective" (p. 57). He acknowledges, "In several of the countries many of the housing units in the Bank-financed projects had in fact been occupied by families with incomes higher than originally intended by the Bank. In some instances . . . it reflected a deliberate decision by the government" (p. 193). He notes that "World Bank resources could free recipient-country resources for the pursuit of other projects" (p. 216). When the Bank financed $23 million for a rural development project and $23.5 million for educational development, Ayres (p. 217) notes that

> the Brazilian government had $46.5 million to spend on other, including non-developmental concerns. Seen in this light, Bank

---

[4] Sowell (1983, p. 239) makes the telling point that the very use of the term "foreign aid" prejudges the result. It "is no more justified *a priori* than calling it 'foreign hindrance.' Whether it is an aid or hindrance is a question of fact in each specific case. Calling the transferring organizations 'development agencies' is likewise begging the central question. They are in fact transfer or donor organizations, and the actual effect of their donations is exactly what is at issue."

> resources financed not only the projects that had been appraised and approved but also projects, perhaps perverse ones, that had not. Even the approved projects may have entailed side benefits going not to the poor but to those allied with the political regime.

And finally, Ayres acknowledges that some bank officials "admitted that they cooked up the evidence" (p. 108).

Implications for Third World Development

Market reforms would do much to stimulate development, but as long as the rulers are the beneficiaries of existing policies there is little hope for dramatic change. By neutralizing the effects of these harmful policies and by providing the elites with a vested interest in their own subjects' poverty, ODA works to perpetuate poverty. Since the bulk of ODA emanates from Western nations, the West has some leverage over recipient countries. The *negative* actions of phasing out "Food for Peace" and other bilateral aid programs, and terminating commitments to such multilateral transfer agencies as the IDA, could induce rulers in at least a few Third World nations to introduce market reforms.

For example, in the 1960s President Julius Nyerere introduced socialist measures in Tanzania. These included extensive nationalizations and the forcible relocation of 13 million people into socialist villages known as *ujamas*. This experiment, said *Time*, in a masterpiece of understatement, went "awry." Production of export crops declined by 40 percent and the manufacturing sector is now operating at less than 20 percent of capacity. Worker output declined by 50 percent and nearly 150 of the 300 companies nationalized by the government were bankrupt by 1975 ("A Continent . . ." 1984, p. 37; Sowell 1983, p. 240). Tanzania receives more ODA per capita than any other nation in the world, and it is doubtful that in its absence what *Time* called the "well-intentioned experiment" could be continued.

The beginning of Taiwan's "economic miracle" coincided with the termination of U.S. economic aid to that country in 1965. The cessation forced Taiwan to abandon its inefficient import substitution policies and move in the direction of free trade (Krauss 1983, p. 159). While one should not be overly optimistic, it is possible that the elimination of ODA, except for temporary emergencies, could trigger other "economic miracles."

Population Control and Growth

Integrally related to the issue of resource availability and economic development is population control. "Overpopulation" only makes sense relative to the "carrying capacity" of a society, which is determined by the economic output of that society. While the nations of the West have more or less achieved population stabilization, the *Global 2000 Report* projects that over 90 percent of the world population growth between 1980 and 2000 will occur in the Third World. It may appear that rapid population growth in an area of the world that is unable to feed itself now is a recipe for disaster, and that only drastic measures to reduce this growth can avert a calamity. This is a staple of Neomalthusian thought, and in order to understand its flaws population growth must be seen in a historical context.

Children are nearly always seen as an economic burden. Historically, however, the reverse was true. In nonindustrialized agrarian societies the cost of rearing children was minimal, and by the time they were five or six they were working in the fields or at other jobs and more than "paying their way." Hence, the decision to have large families was quite rational: children were economic assets; they supported their parents, especially in their old age. The skills demanded in a modern, industrial society often require years of formal training. Hence, as society modernized and children began to be sent to school rather than to the fields, the flow of wealth reversed itself: parents supported their children. This provided a powerful incentive to have fewer children.

There is a related factor. The overwhelming bulk of human history has been characterized by short life expectancies and high mortality rates. High fertility rates were necessary to compensate for these factors. Thanks to increasing productivity and advances in medicine, nutrition, and hygiene, the infant mortality rate declined dramatically in the West during the Industrial Revolution. As life grew more secure people realized that high birth rates were neither necessary nor desirable. For both of these reasons rapid population growth rates proved to be a temporary phenomenon in the West.

The important fact is that this transition from high to low mortality and birth rates is not restricted to the West. It has occurred in all industrialized societies, including such non-Western countries as Japan, Hong Kong, Singapore, Malaysia, Puerto Rico, and Taiwan. And despite individual exceptions, it is occurring throughout the Third World. Between 1950 and 1970 life expectancy in the Third World increased by 58 percent; the birth rate declined by 7 to 8 percent (Coale 1974, p. 51). More recent data show that this adjustment is occurring more rapidly, with fertility rates declining by 16 percent between 1960 and 1980 (Mauldin 1982, p. 255). The world population growth rate peaked at about 2.5 percent per year in the mid-1960s. By 1980 it had declined to 1.7 percent and such forecasters as Kahn (1983, pp. 287–88) expect the decline to continue for the rest of the century. In light of these trends the 1970 UN estimate of a world population of 7.5 billion by the year 2000 has been revised downward to 5.5 billion.

Freedman and Berelson (1974, pp. 31–39) plotted population growth rates over a 16,000 year period, from 8000 B.C. to 8000 A.D. They found that between 8000 B.C. and 1775 A.D. world population grew by no more than 0.1 percent per year. It was only during the mid- to late 18th century that the rate began to climb, peaking during the 1960s. They estimate that the transition from high to low fertility and mortality rates will be completed by about 2150 A.D., when population growth rates will have returned to their pretransition level of about 0.1 percent.

Finally, it should be noted that simply in terms of space there is no population explosion. The world's population density is a mere 75 people per square mile. Marchetti recently calculated that the world could *comfortably* support a trillion people, or 2,500 times the current population, without environmental damage or resource depletion (Louw 1985, p. 15). Given population trends, it is unlikely that this figure will ever be reached. Although everyone knows that we are in the midst of a population explosion, the fact is that the world is nowhere near being overpopulated. Moreover, the scare stories, which are based on the assumption of continued exponential population growth are, as Kahn (1976, p. 30) put it, "at best a rather naive extrapolation of an unusual human experience into the indefinite future without a real understanding of the dynamic forces involved." "The overpopulation worries and alarmist exhortations of the 1960s and 1970s," he wrote, "may well be regarded as an amusing episode in human history."

Implications for Population Control

One cannot deal intelligently with the population issue without also considering economics. The two are often treated separately and result in policies that are limited to artificial methods to reduce the number of births. Since these methods typically ignore the valid reasons for having large families in the Third World, they must often include a large element of coercion and even outright brutality. Such was the case in India during the mid-1970s, when 8.2 million people were involuntarily sterilized. It is also the case in China today, with the ruthless enforcement of the "single-child quota."[5]

But such measures are unnecessary. Shifting the focus from reducing the number of births to stimulating economic development has several important ramifications. First, economic development relieves population pressures by increasing the "carrying capacity" of a given society. Second, by providing a degree of economic security and contributing to a reduction in infant mortality, economic development relieves much of the economic and social necessity for large families. And third, by reversing the flow of wealth it provides a natural, uncoerced incentive for reduced family size.

Since a rise in the standard of living has been historically correlated with declining birth rates, the most efficient and humane birth control program is economic development. And since economic development is closely correlated with a policy of laissez faire, governments that are serious about easing population pressures should adopt a policy of reducing government controls and increasing the protection of private property.

One cautionary note is in order. Since the rulers tend to be the chief beneficiaries of the existing interventionist policies it is unlikely that these policies will be ended. But, here again, there are external "family planning" agencies channeling millions of dollars, annually, to the Third World. These include the UN's Fund for Population Activities and the U.S. Agency for International Development. The termination, or at least reduction, of such aid—together with the reduction or termination of ODA—could be significant enough to induce market-oriented reforms in at least some countries.

Conclusion

According to the Neomalthusians, the world's population is approaching or exceeding the natural capacity of the earth to provide for all of us. The earth, as Loretta McLaughlin (1982) put it, "is headed for the breaking point." The data belie this. Population densities throughout much of the world are quite low. The rate of growth in the world's population has been declining since the 1960s. Per capita food output has been rising, and stocks of resources have been increasing.

---

[5] Fox Butterfield, former head of the *New York Times* Peking bureau (in the *Congressional Record,* 1984), says the "street committees" now have the right to decide "which couples in the neighborhood may have children." He quotes one committee member describing the committee's authority: "We assign a person to keep track of each woman's menstrual cycle. If someone misses her period and isn't scheduled to have a baby, we tell her to have an abortion." According to a local official in charge of administering the policy: "Controlling the population is our aim. . . . Our job is to finish the baby in the stomach. So when you have got rid of the child there is one less person." Yet, China's population density is relatively low; its population growth rate has consistently been below the world's average, and it has more than adequate raw materials and arable land. Its problem is not too many people but rather too little output. And output is low due to the economic strangulation that has occurred since 1949.

Unfortunately, much of the world continues to suffer from malnutrition if not outright starvation. But the principal cause is that throughout much of the world rulers have implemented policies that benefit a tiny elite but harm everyone else. Such policies not only perpetuate poverty but, by keeping the "carrying capacities" below what they would otherwise be, they create "population pressures."

Thus, poverty, malnutrition, and overpopulation are a result not of the natural depletion of the earth's capacity to provide but of policies that artifically impede the efficient use of the earth's endowments. What is sorely needed is for nations to embrace what Adam Smith termed the "natural system of liberty."

References

Ashton, T. S. *The Industrial Revolution, 1760–1830.* London: Oxford University Press, 1960.

Ayres, R. L. *Banking on the Poor.* Cambridge, Mass.: MIT Press, 1983.

Barnett, Harold J., and Morse, Chandler. *Scarcity and Growth.* Baltimore: Johns Hopkins University Press, 1963.

Barnett, Harold J. "Scarcity and Growth Revisited." In *Scarcity and Growth Reconsidered,* pp. 163–217. Edited by V. Kerry Smith. Washington, D.C.: Resources for the Future, 1979.

Barnett, Harold J. et al. "Global Trends in Non-Fuel Minerals." In Simon and Kahn 1984, pp. 316–38.

Barnet, Richard. "No More Room in the Lifeboats." *New York Times Magazine,* 16 April 1978, pp. 32–38.

Barnet, Richard. *The Lean Years: World Resources and the Politics of Scarcity.* New York: Simon and Schuster, 1980.

Bauer, Peter. *Dissent on Development.* Cambridge: Harvard University Press, 1972.

Bauer, Peter. *Equality, the Third World and Economic Delusion.* Cambridge: Harvard University Press, 1981.

Bauer, Peter, and Yamey, Basil. "Foreign Aid: What is at Stake?" In *The Third World,* pp. 115–35. Edited by W. Scott Thompson. San Francisco: Institute for Contemporary Studies, 1983.

Beckerman, Wilfred. *Two Cheers for the Affluent Society.* New York: St. Martin's Press, 1975.

Beckermann, Petr. "Solar Energy and Other 'Alternative' Energy Sources." In Simon and Kahn 1984, pp. 415–27.

Berry, Sara. "Economic Change in Contemporary Africa." In *Africa,* pp. 261–79. Edited by Phyllis Martin and Patrick O'Meara. Bloomington: Indiana University Press, 1977.

Buer, Mabel C. *Health, Wealth and Population in the Early Days of the Industrial Revolution.* New York: Fertig, 1968.

Coale, Ansley. "The History of the Human Population." *Scientific American,* September 1974, pp. 41–52.

Commoner, Barry. *The Poverty of Power.* New York: Knopf, 1976.

Congressional Record. House of Representatives. Vol. 130, no. 59, 9 May 1984.

"A Continent Gone Wrong." *Time,* 16 January 1984, pp. 26–39.

Cook, Earl. "Limits to the Exploitation of Nonrenewable Resources." *Science,* February 1976, pp. 677–82.

Daly, Herman. "The Economics of the Steady State." *American Economic Review* 64 (May 1974): 15–21.

Daly, Herman. "Entropy, Growth, and the Political Economy of Scarcity." In *Scarcity and Growth Reconsidered,* pp. 67–94. Edited by V. Kerry Smith. Washington, D.C., Resources for the Future, 1979.

Eberstadt, Nick. "Hunger and Ideology." *Commentary,* July 1981, pp. 40–49.

Ehrlich, Paul. *The Population Bomb.* New York: Ballantine, 1968.

Ehrlich, Paul. "Eco-Catastrophe." *Ramparts,* September 1969, pp. 24–27.

Ehrlich, Paul, and Ehrlich, Anne. *The End of Affluence.* New York: Ballantine, 1974.

Freedman, Ronald, and Berelson, Bernard. "The Human Population." *Scientific American,* September 1974, pp. 31–39.

Friedman, Milton. "Foreign Economic Aid: Means and Objectives." *Yale Review* (June 1948): 500–16.

Georgescu-Roegen, Nicholas. *The Entropy Law and the Economic Process.* Cambridge: Harvard University Press, 1971.

Georgescu-Roegen, Nicholas. "The Steady State and Ecological Salvation. *Bio-Science,* April 1977, pp. 266–70.

Goeller, H. E. and Weinberg, Alvin M. "The Age of Substitutability." *Science,* February 1976, pp. 683–99.

Goldman, Marshall. *USSR in Crisis.* New York: Norton, 1983.

Hardin, Garrett. "The Tragedy of the Commons." *Science,* December 1968, pp. 1243–48.

Hopper, W. David. "The Development of Agriculture in Developing Countries." *Scientific American,* September 1976, pp. 137–44.

Johnson, D. Gale. "World Food and Agriculture." In Simon and Kahn 1984, pp. 67–112.

Kahn, Herman et al. *The Next 200 Years.* New York: William Morrow, 1976.

Kahn, Herman, and Schneider, Ernest. "Globaloney 2000." In *World Politics Debated,* pp. 383–93. Edited by Herbert Levine. New York: McGraw-Hill, 1983.

Kirzner, Israel. *Competition and Entrepreneurship.* Chicago: University of Chicago Press, 1973.

Krauss, Melvyn. *Development Without Aid: Growth, Poverty and Government.* New York: McGraw-Hill, 1983.

Louw, Leon. "A Critical Review of Prevailing Energy Predictions and Policies." Address to the Thirteenth Annual ALESEC Economic Congress on

"Energy: A Factor in Economic Development—A South African Perspective." 1985.

Mauldin, W. Parker. "Fertility and Population Growth." In *International Encyclopedia of Population*, pp. 252–58. Edited by John A. Ross. New York: Free Press, 1982.

McLaughlin, Loretta. "Earth Headed for the Breaking Point." *Chicago Tribune*, 19 September 1982.

Meadows, Donella et al. *The Limits to Growth*. New York: Universe Books, 1972.

Mehta, Ved. *The New India*. New York: Penguin, 1978.

National Commission on Air Quality. *To Breathe Clean Air*. Washington, D.C.: Government Printing Office, 1981.

Nordhaus, William D. "Resources as a Constraint on Growth." *American Economic Review* 64 (May 1974): 22–26.

Osborne, David. "America's Plentiful Energy Resource." *Atlantic Monthly*, March 1984, pp. 86–102.

Osterfeld, David. "Assessing the New International Order: Prospects for Third World Development." *Journal of Social, Political and Economic Studies* (Spring/Summer 1982): 3–26.

Osterfeld, David. "The Increasing Abundance of Resources." *The Freeman*, June 1984, pp. 360–77.

Paddock, William, and Paddock, Paul. *Famine—1975! America's Decision: Who Will Survive?* Boston: Little, Brown, 1967.

Perlman, Lewis. *The Global Mind*. New York: Mason/Charter, 1976.

Ridker, Ronald, and Cecelski, Elizabeth. "Resources and Population." In *International Encyclopedia of Population*, pp. 590–600. Edited by John Ross. New York: Free Press, 1982.

Ritter, Malcolm. "Biotechnology breaking the 'species barrier'." *Chicago Tribune*, 5 August 1984.

Simon, Julian. "Resources, Population, Environment: An Oversupply of False Bad News." *Science*, July 1980, pp. 1431–37.

Simon, Julian. "Global Confusion, 1980: A Hard Look at the *Global 2000 Report*." *The Public Interest* (Winter 1981a): 3–20.

Simon, Julian. *The Ultimate Resource*. Princeton: Princeton University Press, 1981b.

Simon, Julian, and Kahn, Herman eds. *The Resourceful Earth*. New York: Blackwell, 1984.

Solow, Robert. "The Economics of Resources or the Resources of Economic." *American Economic Review* 64 (May 1974):1–14.

Sowell, Thomas. *Knowledge and Decisions*. New York: Basic Books, 1980.

Sowell, Thomas. *The Economics and Politics of Race*. New York: William Morrow, 1983.

Stroup, Richard, and Baden, John. *Natural Resources: Bureaucratic Myths and Environmental Management*. San Francisco: Pacific Institute for Public Policy Research, 1983.

U.N. Food and Agricultural Organization. *The Fourth World Food Survey*. Rome: United Nations, 1977.

U.S. Council on Environmental Quality. *Environmental Trends*. Washington, D.C.: Government Printing Office, 1981.

U.S. Council on Environmental Quality and the Department of State. *The Global 2000 Report to the President*. New York: Penguin, 1982.

U.S. Department of Commerce. *Statistical Abstract of the United States*. Various Years. Washington, D.C.: Government Printing Office.

Wiser, William, and Wiser, Charlotte. *Behind Mud Walls*. Berkeley: University of California Press, 1971.

World Bank. *World Development Report 1981*. Washington, D.C.: World Bank, 1981.

A PUBLIC CHOICE VIEW OF THE MINIMUM WAGE

Thomas Rustici

> Why, when the economist gives advice to his society, is he so often cooly ignored? He never ceases to preach free trade . . . and protectionism is growing in the United States. He deplores the perverse effects of minimum wage laws, and the legal minimum is regularly raised each 3 or 5 years. He brands usury laws as a medieval superstition, but no state hurries to repeal its law.
>
> —George Stigler

I. Introduction

Much of public policy is allegedly based on the implications of economic theory. However, economic analysis of government policy is often disregarded for political reasons. The minimum wage law is one such example. Every politician openly deplores the spectacle of double-digit teenage unemployment pervading modern society. But, when economists claim that scientific proof, a priori and empirical, dictates that minimum wage laws cause such a regretful outcome, their statements generally fall on deaf congressional ears. Economists too often assume that policymakers are interested in obtaining all the existing economic knowledge before deciding on a specific policy course. This view of the policy-formation process, however, is naive. In framing economic policy politicians will pay some attention to economists' advice, but such advice always will be rejected when it conflicts with the political reality of winning votes. At bottom, what is important in analyzing the course of economic policy is an understanding of the public choice aspects of the decision-making

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is a graduate student in economics at the Center for the Study of Market Processes, George Mason University. An earlier version of this paper was awarded the 1984 Excellence in Liberty Prize by the Institute for Humane Studies. The author wishes to thank Thomas J. DiLorenzo for his helpful comments.

process in government. In such an undertaking, the incentive structure confronting governmental decision makers must be carefully examined.

This paper therefore takes a public-choice perspective of minimum wage laws, explaining their persistent political support by examining the underlying incentive structure. Although the unemployment effects of such laws have been public knowledge for decades, and despite nearly unanimous opposition to them within the economics profession, they are now as solidly in place as ever before. Some conflicting conclusions can be drawn from this paradox: (1) the majority of economists are completely wrong on this issue, and have been for decades; or (2) politicians who claim opposition to unemployment and then call on economists to testify about the minimum wage before voting to increase it are not really interested in what economists have to say; or (3) economists have not been able to convince the average voter that it is in his interest to reject the minimum wage.

If the first conclusion is true, economists need to reevaluate the quality of their evidence on the economic effects of the minimum wage. The second conclusion would indicate that economists have very little influence over the direction of public policy disputes in advising politicians, who are constrained by the incentive structure of the "political market." The final reason for the paradox would point to a needed change in the direction of emphasis taken by economists in dealing with the minimum wage.

The relevant question for this paper, therefore, is the following: If the minimum wage law has consequences deemed undesirable by those who implement it, and if evidence accumulated over the decades conclusively proves these consequences to be inevitable features of the law, then why does it continue to survive? To answer this question, the paper begins with an examination of the economic effects of the minimum wage in section II. The sources of political support for the minimum wage law are then explored in section III. Section IV looks at the minimum wage issue from a North-South perspective and discusses the beneficiaries and victims of the minimum wage. Given the political environment of past minimum wage legislation, section V considers the prospects of real reform, and section VI concludes with the proviso that economists ought to direct their reform efforts at educating the public on the adverse effects of the minimum wage rather than trying to influence legislators directly.

II. Economic Effects of the Minimum Wage

Economic analysis has demonstrated few things as clearly as the effects of the minimum wage law. It is well known that the minimum wage creates unemployment among the least skilled workers by raising wage rates above free market levels. Eight major effects of the minimum wage can be discussed: unemployment effects, employment effects in uncovered sectors of the economy, reduction in nonwage benefits, labor substitution effects, capital substitution effects, racial discrimination in hiring practices, human capital development, and distortion of the market process with respect to comparative advantage. Although the minimum wage has other effects, such as a reduction in hours of employment, these eight effects are the most significant ones for this paper.

Unemployment Effects

The first federal minimum wage laws were established under the provisions of the National Recovery Administration (NRA). The National Industrial Recovery Act, which became law on 16 June 1933, established industrial minimum wages for 515 classes of labor. Over 90 percent of the minimum wages were set at between 30 and 40 cents per hour.[1] Early empirical evidence attests to the unemployment effects of the minimum wage. Using the estimates of C. F. Roos, who was the director of research at the NRA, Benjamin Anderson states: "Roos estimates that, by reason of the minimum wage provisions of the codes, about 500,000 Negro workers were on relief in 1934. Roos adds that a minimum wage definitely causes the displacement of the young, inexperienced worker and the old worker."[2]

On 27 May 1935 the Supreme Court declared the NRA unconstitutional, burying the minimum wage codes with it. The minimum wage law reappeared at a later date, however, with the support of the Supreme Court. In what became the precedent for the constitutionality of future minimum wage legislation, the Court upheld the Washington State minimum wage law on 29 March 1937 in *West Coast Hotel v. Parrish.*[3] This declaration gave the Roosevelt administration and Labor Secretary Frances Perkins the green light to reestablish the federal minimum wage, which was achieved on 25 June 1938 when President Roosevelt signed into law the Fair Labor Standards Act (FLSA).

The FLSA included legislation affecting work-age requirements, the length of the workweek, pay rates for overtime work, as well as the national minimum wage provision. The law established minimum wage rates of 25 cents per hour the first year, 30 cents per hour for the next six years, and 40 cents per hour after seven years. The penalty for noncompliance was severe: violators faced a $10,000 fine, six months imprisonment, or both. In addition, an aggrieved employee could sue his employer for twice the difference between the statutory wage rate and his actual pay.[4]

---

[1] Leverett Lyon, et al. *The National Recovery Administration: An Analysis and Appraisal* (New York: Da Capo Press, 1972), pp. 318–19.

[2] Benjamin M. Anderson, *Economics and the Public Welfare: A Financial and Economic History of the United States, 1914–1946* (Indianapolis: Liberty Press, 1979), p. 336.

[3] Jonathan Grossman, "Fair Labor Standards Act of 1938: Maximum Struggle for a Minimum Wage," *Monthly Labor Review* 101 (June 1978): 23.

With the passage of the FLSA, it became inevitable that major dislocations would result in labor markets, primarily those for low-skilled and low-wage workers. Although the act affected occupations covering only one-fifth of the labor force,[5] leaving a large uncovered sector to minimize the disemployment effects, the minimum wage was still extremely counterproductive. The Labor Department admitted that the new minimum wage had a disemployment effect, and one historian sympathetic to the minimum wage was forced to concede that "[t]he Department of Labor estimated that the 25-cents-an-hour minimum wage caused about 30,000 to 50,000 to lose their job. About 90% of these were in southern industries such as bagging, pecan shelling, and tobacco stemming."[6]

These estimates seriously understate the actual magnitude of the damage. Since only 300,000 workers received an increase as a result of the minimum wage,[7] estimates of 30,000–50,000 lost jobs reveal that 10–13 percent of those covered by the law lost their jobs. But it is highly dubious that only 30,000–50,000 low-wage earners lost their jobs in the entire country; that many unemployed could have been found in the state of Texas alone, where labor authorities saw devastation wrought via the minimum wage on the pecan trade. The *New York Times* reported the following on 24 October 1938:

> Information received today by State labor authorities indicated that more than 40,000 employees of the pecan nut shelling plants in Texas would be thrown out of work tomorrow by the closing down of that industry, due to the new Wages and Hours Law. In San Antonio, sixty plants, employing ten thousand men and women, mostly Mexicans, will close. . . . Plant owners assert that they cannot remain in business and pay the minimum wage of 25 cents an hour with a maximum working week of forty-four hours. Many garment factories in Texas will also close.[8]

---

[4] "Wage and Hours Law," *New York Times*, 24 October 1938, p. 2.

[5] Grossman, "Fair Labor Standards Act," p. 29.

[6] Ibid., p. 28.

[7] Ibid., p. 29.

[8] "Report 40,000 Jobs Lost," *New York Times*, 24 October 1938, p. 2.

It can reasonably be deduced that even if the Texas estimates had been wildly inaccurate, the national unemployment effect would still have exceeded the Department of Labor's estimates.

The greatest damage, however, did not come in Texas or in any other southern state, but in Puerto Rico. Since a minimum wage law has its greatest unemployment effect on low-wage earners, and since larger proportions of workers in poorer regions such as Puerto Rico tend to be at the lower end of the wage scale, Puerto Rico was disproportionately hard-hit. Subject to the same national 25-cents-per-hour rate as workers on the mainland, Puerto Rican workers suffered much more hardship from the minimum wage law. According to Anderson:

> It was thought by many that, in the first year, the provision would not affect many industries outside the South, though the framers of the law apparently forgot about Puerto Rico, and very grave disturbances came in that island. . . . Immense unemployment resulted there through sheer inability of important industries to pay the 25 cents an hour.[9]

Simon Rottenberg likewise points out the tragic position in which Puerto Rico was placed by the enactment of the minimum wage:

> When the Congress established a minimum wage of 25 cents per hour in 1938, the average hourly wage in the U.S. was 62.7 cents. . . . It resulted in a mandatory increase for only some 300,000 workers out of a labor force of more than 54 million. In Puerto Rico, in contrast . . . the new Federal minimum far exceeded the prevailing average hourly wage of the major portion of Puerto Rican workers. If a continuing serious attempt at enforcement . . . had been made, it would have meant literal economic chaos for the island's economy.[10]

On 24 October 1938, a special cable from Puerto Rico was printed in the *New York Times* detailing the effects of the new wage law on Puerto Rican workers:

> Wage payrolls estimated at approximately $1,000,000 monthly by the Chamber of Commerce will end tomorrow with the application of the Fair Labor Standards Act. Both labor and employers appear to be united in the position that the law applied to Puerto Rico ends employment for approximately 120,000 persons. It is also believed to terminate prospects for any possible further industrialization.
>
> Recent conferences in Washington between Administrator Elmer F. Andrews and island representatives led to the conclusion that

---

[9] Anderson, *Economics and the Public Welfare*, p. 458.

[10] Simon Rottenberg, "Minimum Wages in Puerto Rico," in *Economics of Legal Minimum Wages*, edited by Simon Rottenberg (Washington, D.C.: American Enterprise Institute, 1981), p. 330.

> there was not any means of modifying the effects of the law in the island short of a Congressional amendment. With the unemployment resulting from the law's enforcement, the Chamber of Commerce study shows that the 230,000 normally unemployed, out of a total of 650,000 employables there, will be increased to 350,000, or more than half of Labor's ranks in the island. . . .
>
> Much of labor is more than anxious to continue at the prevailing rates on the theory that half a loaf is better than none. Labor Commissioner Prudenceo Rivera Martinez, in a published statement today said: "The medicine is too strong for the patient."[11]

Puerto Rico's unemployment rate rose sharply due to the Fair Labor Standards Act. Tens of thousands lost their jobs in such industries as cigar and cigarette manufacturing, which all but disappeared.[12] The needlework industry, which employed over 40,000 workers in 1935, stagnated after the new act took effect. The value of needle-trade exports fell from over $20 million before the act in 1937 to barely $5 million in 1940.[13] Rafael Pico describes the plight of the needlework industry after the act took effect:

> An industry which had been paying three or four cents per hour for work in homes could obviously not survive under the drastic mandatory increase. The effects of the law . . . were catastrophic. . . . The industry would have disappeared entirely if a legislative appeal for special legislation on minimum wages had not been sent to the Congress of the U.S.[14]

After two years of economic disruption in Puerto Rico, Congress amended the minimum wage provisions.[15] The minimum wage was reduced to 12.5 cents per hour, but it was too late for many industries and for thousands of low-wage earners employed by them, who suddenly found unemployment the price they had to pay for the minimum wage.

In sum, the tragedy of the minimum wage laws during the NRA and the FLSA was not just textbook-theorizing by academic economists, but real-world disaster for the thousands who became the victims of the law. But these destructive effects have not caused the law to be repealed; to the contrary, it has been expanded in coverage and increased in amount.

---

[11] "Puerto Rico Hurt by Wage Hour Law," *New York Times*, 24 October 1938, p. 2.

[12] Rafael Pico, *The Geography of Puerto Rico* (Chicago: Aldine Publishing, 1974), p. 293.

[13] Ibid., p. 294.

[14] Ibid.

[15] Rottenberg, "Minimum Wages in Puerto Rico," p. 333.

Meanwhile, as can be seen from Table 1, the evidence for the unemployment effects of the minimum wage continues to mount. Many empirical studies since the early 1950s—from early research by Marshall Colberg and Yale Brozen to more recent work by Jacob Mincer and James Ragan—have validated the predictions of economic theory regarding the unemployment effects of the minimum wage law. In virtually every case it was found that the net employment effects and labor-force participation rates were negatively related to changes in the minimum wage. In the face of 50 years of evidence, the question is no longer *if* the minimum wage law creates unemployment, but *how much* current or future increases in the minimum wage will adversely affect the labor market.

Employment in Uncovered Sectors

The labor market can be divided into two sectors: that covered by the minimum wage law, and that not covered. In a partially covered market, the effects of the minimum wage are somewhat disguised. Increasing it disemploys workers in the covered sector, prompting them to search for work in the uncovered sector if they are trainable and mobile. This then drives down the wage rate in the uncovered sector, making it lower than it otherwise would have been. Since perfect knowledge and flexibility is not observed in real-world labor markets, substantial unemployment can occur during the transition period.

Employees in the covered sector who do not lose their jobs get a wage-rate increase through the higher minimum wage. But this comes only at the expense of (1) the disemployed workers who lose their jobs and suffer unemployment during the transition to employment in the uncovered sector, and (2) everyone in the uncovered sector, as their wage rate falls due to the influx of unemployed workers from the covered sector. While increasing the incomes of some low-wage earners, increasing the minimum wage tends to make the lowest wage earners in the uncovered sector even poorer than they otherwise would have been.

Yale Brozen has found that the uncovered household sector served to absorb the minimum wage–induced disemployed in the past.[16] But the "safety valve" of the uncovered portion of the economy is rapidly vanishing with the continual elimination of various

---

[16] Yale Brozen, "Minimum Wage Rates and Household Workers," *Journal of Law and Economics* 5 (October 1962): 103–10.

TABLE 1

UNEMPLOYMENT EFFECTS OF THE MINIMUM WAGE: A SURVEY

Author(s)	Source	Empirical Findings & Conclusions
Marshall Colberg	"Minimum Wage Effects on Florida's Economic Development," *Journal of Law and Economics* (October 1962)	The response to the increase in the minimum wage from $0.75 per hour to $1.00 per hour resulted in a 15.2 percent reduction in man-hours of employment for production workers in low-wage counties during January–April 1956. Weeks of unemployment compensation paid for low-wage counties increased 67.8 percent during the same period.
Yale Brozen	"Minimum Wage Rates and Household Workers," *Journal of Law and Economics* (April 1962)	Household employment, which usually moves contracyclically, failed to do so when the minimum wage was increased. The uncovered sector of the household labor market absorbed those workers who became displaced by the higher minimum wage.
Arthur Burns	*The Management of Prosperity* (New York: Columbia University Press, 1966)	An increase of $0.25 per hour in the minimum wage would raise the unemployment rate for nonwhite teenagers by 8 percent.
Yale Brozen	"The Effects of the Statutory Minimum Wage Increases on Teen-Age Unemployment," *Journal of Law and Economics* (April 1969)	Successive increases in the minimum wage were directly followed by increases in the teenage unemployment rate.
Finis Welch	"Minimum Wage Legislation in the United States," *Economic Inquiry* (September 1974)	The minimum wage reduced employment and heightened vulnerability of teenage employment over the course of the business cycle.

Edward Gramlich	"Impact of Minimum Wage on Other Wages, Employment, and Family Income," *Brookings Papers on Economic Activity* 2 (1976)	The minimum wage creates unemployment, while the bulk of increased benefits tend to go to higher income families.
Jacob Mincer	"Unemployment Effects of Minimum Wages," *Journal of Political Economy* (August 1976)	The net effect of the minimum wage on labor force participation is negative, with the largest negative effect observed for nonwhite teenagers. The net employment effect is also negative, with the largest disemployment occurring in the nonwhite teenage group.
U.S. Chamber of Commerce	*The Congressional Record*, 6 October 1977, p. 32708	The estimated impact of the proposed increases in the minimum wage from $2.30 to $3.15 in three steps through January 1980 would lead to a national 1,977,000 job loss, and a 4.1 percent increase in labor costs.
James Ragan	"Minimum Wages and the Youth Labor Market," *Review of Economics and Statistics* (May 1977)	Had the 1966 amendment to increase the minimum wage not occurred, aggregate youth employment would have been 225,000 higher in 1972, and the youth unemployment rate 3.8 percent lower.
William Beranek	"The Illegal Alien Work Force, Demand for Unskilled Labor, and the Minimum Wage," *Journal of Labor Research* (Winter 1982)	"The abolition of minimum wage laws and alternative income support programs in 1978 could have increased employment levels of domestic unskilled workers from roughly 500,000 to 2 million."[a]
Robert H. Meyer & David A. Wise	"The Effects of the Minimum Wage on the Employment and Earnings of Youth," *Journal of Labor Economics* 1, no. 1 (1983)	"Among black youth with market wage rates below the minimum, nonemployment is increased from 25.3% without the minimum to 43.4% with the minimum."[a]

---

[a] Direct quotation from the author(s).

exemptions.[17] Because of this trend we can expect to see the level of structural unemployment increase with escalation of the minimum wage.[18]

Nonwage Benefits

Wage rates are not the only costs associated with the employment of workers by firms. The effective labor cost a firm incurs is usually a package of pecuniary and nonpecuniary benefits. As such, contends Richard McKenzie,

> employers can be expected to respond to a minimum wage law by cutting back or eliminating altogether those fringe benefits and conditions of work, like the company parties, that increase the supply of labor but which do not affect the productivity of labor. By reducing such non-money benefits of employment, the employer reduces his labor costs from what they otherwise would have been and loses nothing in the way of reduced labor productivity."[19]

If one takes the view that employees desire both pecuniary and nonpecuniary income, then anything forcing them to accept another mix of benefits would clearly make them worse off. For example, suppose worker A desires his income in the form of $3.00 per hour in wages, an air-conditioned workplace, carpeted floors, safety precautions, and stereo music. If he is *forced* by the minimum wage law to accept $3.25 per hour and fewer nonpecuniary benefits, he is worse off than at the preminimum wage and the *higher* level of nonpecuniary income. A priori, the enactment of minimum wage laws must place the worker and employer in a less-than-optimal state. Thus it may not be the case that only unemployed workers suffer from the minimum wage; even workers who receive a higher wage and retain employment may be net losers if their nonpecuniary benefits are reduced.

Labor Substitution Effects

The economic world is characterized by a plethora of substitutes. In the labor market low-skill, low-wage earners are substitutes for high-skill, high-wage earners. As Walter Williams points out:

> Suppose a fence can be produced by using either one high skilled worker or by using three low skilled workers. If the wage of high skilled workers is $38 per day, and that of a low skilled worker is $13 per day, the firm employs the high skilled worker because costs

---

[17] Finis Welch, "Minimum Wage Legislation in the United States," *Economic Inquiry* 12 (September 1974): 286.

[18] Brozen, "Minimum Wage Rates and Household Workers," pp. 107–08.

[19] Richard McKenzie, "The Labor Market Effects of Minimum Wage Laws: A New Perspective," *Journal of Labor Research* 1 (Fall 1980): 258–59.

> would be less and profits higher ($38 versus $39). The high skilled worker would soon recognize that one of the ways to increase his wealth would be to advocate a minimum wage of, say, $20 per day in the fencing industry. . . . After enactment of the minimum wage laws, the high skilled worker can now demand any wage up to $60 per day . . . and retain employment. Prior to the enactment of the minimum wage of $20 per day, a demand of $60 per day would have cost the high skilled worker his job. Thus the effect of the minimum wage is to price the high skilled worker's competition out of the market.[20]

Labor competes against labor, not against management. Since low-skill labor competes with high-skill labor, the minimum wage works against the lower-skill, lower-paid worker in favor of higher-paid workers. Hence, the consequences of the law are exactly opposite its alleged purpose.

Capital Substitution Effects

To produce a given quantity of goods, some bundle of inputs is required. The ratio of inputs used to produce the desired output is not fixed by natural law but by the relative prices of inputs, which change continuously with new demand and supply conditions. Based on relative input prices, producers attempt to minimize costs for a given output. Since many inputs are substitutes for one another in the production process, a given output can be achieved by increasing the use of one and diminishing the use of another. The optimal mix will depend on the relative supply and demand for competing substitute inputs.

As a production input, low-skill labor is often in direct competition with highly technical machinery. A Whirlpool dishwasher can be substituted for low-skill manual dishwashers in the dishwashing process, and an automatic elevator can take the place of a nonautomatic elevator and a manual operator. This not to imply that automation "destroys jobs," a common Luddite myth. As Frederic Bastiat explained over a century ago, jobs are obstacles to be overcome.[21] Automation shifts the *kinds* of jobs to be done in society but does not reduce their total number. Low-skill jobs are done away with, but higher-skill jobs are created simultaneously. When the minimum wage raises the cost of employing low-skill workers, it makes the substitute of automated machinery an attractive option.

---

[20] Walter Williams, *The State Against Blacks* (New York: McGraw-Hill, 1982), pp. 44–45.

[21] Frederic Bastiat, *Economic Sophisms* (Irvington-on-Hudson, N.Y.: Foundation for Economic Education, 1946), pp. 16–19.

Racial Discrimination in Hiring Practices

At first glance the connection between the level of racial discrimination in hiring practices and the minimum wage may not seem evident. On closer examination, however, it is apparent that the minimum wage law gives employers strong incentives to exercise their existing racial preferences.[22] The minimum wage burdens minority groups in general and minority teenagers most specifically. Although outright racism has often been blamed as the sole cause of heavy minority teenage unemployment, it is clearly not the only factor. William Keyes informs us that

> In the late 1940's and early 1950's, young blacks had a lower unemployment rate than did whites of the same age group. But after the minimum wage increased significantly, especially in 1961, the black youth unemployment rate has increased to the extent that it is now a multiple of the white youth unemployment rate.[23]

To make the case that racism itself is the cause of the employment and unemployment disparity among blacks and whites, one would have to claim that America was more racially harmonious in the past than it is now. In fact, during the racially hostile times of the early 1900s 71 percent of blacks over nine years of age were employed, as compared with 51 percent for whites.[24] The minimum wage means that employers are not free to decide among low-wage workers on the basis of price differentials; hence, they face fewer disincentives to deciding according to some other (possibly racial) criteria.

To see the racial implications of minimum wage legislation, it is helpful to look at proponents of the law in a country where racial hostility is very strong, South Africa. Since minimum wage laws share characteristics in common with equal pay laws, white racist unions in South Africa continually support both minimum wage and equal-pay-for-equal-work laws for blacks. According to Williams:

> Right-wing white unions in the building trades have complained to the South African government that laws reserving skilled jobs for whites have been broken and should be abandoned in favor of equal pay for equal work laws. . . . The conservative building trades made it clear that they are not motivated by concern for black workers but had come to feel that legal job reservation had been so eroded by government exemptions that it no longer protected the white worker.[25]

---

[22] Walter Williams, "Government Sanctioned Restraints That Reduce the Economic Opportunities for Minorities," *Policy Review* 22 (Fall 1977): 15.

[23] William Keyes, "The Minimum Wage and the Davis Bacon Act: Employment Effects on Minorities and Youth," *Journal of Labor Research* 3 (Fall 1982): 402.

[24] Williams, *State Against Blacks*, p. 41.

[25] Ibid., p. 43.

The reason white trade unions are restless in South Africa is a $1.52-per-hour wage differential between black and white construction workers.[26] Although the owners of the construction firms are white, they cannot afford to restrict employment to whites when blacks are willing to work for $1.52 per hour less. As minimum wages eliminate the wage differential, the cost to employers of hiring workers with the skin color they prefer is reduced. As the cost of discrimination falls, and with all else remaining the same, the law of demand would dictate that more discrimination in employment practices will occur.

Markets frequently respond where they can, even to the obstacles the minimum wage presents minority groups. In fact, during the NRA blacks would frequently be advanced to the higher rank of "executives" in order to receive exemptions from the minimum wage.[27] The free market demands that firms remain color-blind in the conduct of business: profit, not racial preference, is the primary concern of the profit-maximizing firm. Those firms who fail the profit test get driven out of business by those who put prejudice aside to maximize profits. When markets are restricted by such laws as the minimum wage, the prospects for eliminating racial discrimination in hiring practices and the shocking 40–50 percent rate of black teenage unemployment in our cities are bleak.

Human Capital Development

Minimum wage laws restrict the employment of low-skill workers when the wage rate exceeds the workers' marginal productivity. By doing so, the law prevents workers with the least skills from acquiring the marketable skills necessary for increasing their future productivity, that is, it keeps them from receiving on-the-job training.

It is an observable fact, true across ethnic groups, that income rises with age.[28] As human capital accumulates over time, it makes teenagers more valuable to employers than workers with no labor-market experience. But when teenagers are priced out of the labor market by the minimum wage, they lose their first and most crucial opportunity to accumulate the human capital that would make them more valuable to future employers. This stunting reduces their lifetime potential earnings. As Martin Feldstein has commented:

> [F]or the disadvantaged young worker, with few skills and below average education, producing enough to earn the minimum wage

---

[26] Ibid., pp. 43–44.

[27] Lyon, *National Recovery Administration*, p. 339.

[28] U.S., Department of Commerce, Bureau of the Census, *Statistical Abstract of the United States 1982–83*, p. 431.

> is incompatible with the opportunity for adequate on-the-job learning. For this group, the minimum wage implies high short-run unemployment and the chronic poverty of a life of low wage jobs.[29]

Feldstein also finds a significant irony in the minimum wage: "It is unfortunate and ironic that we encourage and subsidize expenditure on formal education while blocking the opportunity for individuals to 'buy' on-the-job training."[30] This is especially hard on teenagers from the poorest minority groups, such as blacks and hispanics—a truly sad state of affairs, since the law is instituted in the name of the poor.

Distortion of the Market Process

Relative prices provide the transmission mechanism by which information is delivered to participants in the market about the underlying relative scarcities of competing factor inputs. They serve as signals for people to substitute relatively less scarce resources for relatively more scarce resources, in many cases without their even being aware of it.[31]

Whenever relative price differentials exist for input substitutes in the production process, entrepreneurs will switch from higher-priced inputs to lower-priced inputs. In a dynamically changing economy, this switching occurs continually. But when prices are not allowed to transmit market information accurately, as in the case of prices artificially controlled by government, then distorted information skews the market and guides it to something clearly less than optimal.[32]

Minimum wages, being such a distortion of the price system, lead to the wrong factor input mix between labor and all other inputs. As a result, industry migrates to locations of greater labor supply more slowly, and labor-intensive industries tend to remain fixed in non-optimal areas, areas with greater labor scarcity. Large labor pools of labor-abundant geographical areas are not tapped because the controlled price of labor conveys the wrong information to all the parties involved. Thus, the existence of price differentials, as knowledge to be transmitted through relative prices, is hidden.[33] The slowdown of industrial migration keeps labor-abundant regions poorer than they otherwise would be because economic growth there is stifled. As

---

[29] Martin Feldstein, "The Economics of the New Unemployment," *The Public Interest*, no. 33 (Fall 1973): 14–15.

[30] Ibid., p. 15.

[31] Thomas Sowell, *Knowledge and Decisions* (New York: Basic Books, 1980), p. 79.

[32] Ibid.

[33] Ibid., pp. 167–68.

Simon Rottenberg explains for the case of Puerto Rico:

> The aggregate effect of all these distortions was that Puerto Rico could be expected to produce fewer goods and services than would have otherwise been produced and that the rate at which insular per capita income rose toward mainland United States income standards could be expected to be dampened. In sum, the minimum wage law could be expected to reduce the rate of improvement in the standard of life of the Puerto Rican people and to intensify poverty in the island.[34]

In summary, the evidence is in on the minimum wage. All eight major effects of the minimum wage examined here make the poor, disadvantaged, or young in society worse off—the alleged beneficiaries turn out to be the law's major victims. So why does a law whose consequences its own designers would officially declare to be "bad" on all counts continue to survive decade after decade? Why, moreover, is it periodically expanded in scope and raised in amount? The answer is to be found by examining the actual beneficiaries of the law and the functioning of the political market.

III. Sources of Political Support for the Minimum Wage

Unions and the Minimum Wage

Unions everywhere support the minimum wage, a fact that could be deduced from the above analysis of the labor substitution effect of the minimum wage law. Unions are labor cartels that attempt to restrict the supply of workers entering given occupations. Since nonunion labor is priced below the cartelized price of union labor, it is an attractive substitute for union workers. Because unionization of *all* potential competition to the cartel is impossible due to the high policing costs that would be involved, unions resort to the minimum wage. By artificially increasing the wage rate of lower-skilled workers—who could substitute for union workers—the minimum wage increases the demand for union workers and hence their wage rates. Thus, government enforcement of a minimum wage allows unions to (1) keep a stronger cartel, raising their income; (2) incur no policing costs; and (3) have the government force low-skilled, nonunion workers out of the labor market.

It is important to note that unions unanimously supported the cartelization created by the NRA codes.[35] When the NRA and the minimum wage provisions were abolished in 1935 by the Supreme

---

[34] Rottenberg, "Minimum Wages in Puerto Rico," p. 329.

[35] Jesse Thomas Carpenter, *Competition and Collective Bargaining in the Needle Trades 1910—1967* (New York: W. F. Humphrey Press, 1972), p. 759.

Court, unions scurried to reestablish wage floors everywhere. The recurrent phenomenon of firms with lower labor costs outcompeting unionized firms presented a serious problem for such labor unions as the International Ladies Garment Workers Union, which faced its fiercest competition from nonunion garment factories in the southern United States and in Puerto Rico. The efforts of major unions to equalize labor costs nationally were futile.[36] A substitute for the defunct NRA codes was needed by the unions and was found in the Fair Labor Standards Act. Former senator Paul Douglas notes the support of the FLSA by prominent union leaders:

> Sidney Hillman, president of the Amalgamated Clothing Workers, an important CIO union, who was in charge of the organization drive in the textile industry, testified, however, that he wanted this power of the Board retained, since he believed that legislation should not stop at an $800 full-time yearly wage. He, moreover, declared that such higher rates would aid rather than hinder collective bargaining by protecting the high wage firms from undue wage-cutting by competitors.[37]

On 19 July 1955, testimony was entered in the *Congressional Record* from the presidents of four major unions concerning the minimum wage in Puerto Rico. The union presidents demanded that the Puerto Rican minimum be increased in spite of the opposition of Governor Luis Munoz-Marin.[38] It is just not credible that the presidents of the Amalgamated Clothing Workers, International Ladies Garment Workers Union, United Hatters, Cap and Millinery Workers, and the Textile Workers Union were all sincerely misguided humanitarians who happened not to understand the economic effects of the law. To the contrary, they knew that Puerto Rican wage differentials spelled trouble for members of their unions.

> The question of the minimum wages to be required in Puerto Rico under the act is of special interest to the four unions [the four unions mentioned above], which have formed the Joint Minimum Wage Committee. . . . Yet, today, the wage gap between Puerto Rico and the mainland is wider, not narrower, than it was in 1949, and Puerto Rico enjoys greater, not less, competitive advantages over the mainland. . . . The lack of proper wage advances in Puerto Rico constitutes an increasingly unfair threat to the mainland. This is a situation which cannot be left unchallenged by your committee and the House of Representatives.[39]

---

[36] Ibid., p. 815.

[37] Paul Douglas, "Fair Labor Standards Act of 1938," *Political Science Quarterly* 53 (December 1938): 501.

[38] *Congressional Record,* 19 July 1955, p. 10977.

[39] Ibid.

In what must go down in history as a classic example of political double-talk, the union presidents said that "[t]he unions joining in this communication are all friends of Puerto Rico. If we were unconcerned about the island's economic welfare we would be urging uniform minimum wages for Puerto Rico and the mainland."[40]

The union presidents were declaring explicitly that they understood the economic effects of the law. They also understood their self-interest, however, and therefore denounced the wage differentials. What is so astounding is that they would first express their knowledge that elimination of the wage differential would harm the island, then demand an increase in the minimum wage, and then claim to be friends of Puerto Rico! Rottenberg sums up the unions' motivation in getting the law applied to Puerto Rico:

> It is clear that their intention is not to improve the conditions of Puerto Rican workers so much as to deprive those workers of employment opportunities by compelling them to offer their services at a high legally defined price. Their interest lies in influencing the spatial distribution of particular kinds of economic activity so that more of the kinds of goods produced by their members will be made on the mainland and less in Puerto Rico. They want to effect the distribution of wealth and income to the advantage of particular sets of mainland workers at the expense of Puerto Rican workers. The minimum wage law is an instrumental tactic employed by unions to achieve that purpose.[41]

Trade unions are highly organized institutions, and they do not take lightly the benefits that accrue to them through minimum wage laws. At their disposal are large numbers of highly organized voting members, as well as considerable financial/political clout.[42] Five of the ten largest political action committees in 1979–80 were unions.[43] The greatest concentration of benefits would, a priori, be thought to accrue to unions in such occupations as apparel, textiles, and agriculture, where direct competition with nonunion workers is greatest. And indeed, these are the unions we actually see consistently lobbying for higher minimum wages.

It is clear that unions have no intention of improving the lot of poorer workers through the minimum wage law, especially those in Puerto Rico and the Virgin Islands. Since there are no official trade barriers between U.S. territories and the mainland, the minimum wage serves as a proxy for tariffs on goods imported from lower-wage territories. It is hard to believe that the supposedly benign intentions of unions are usually accepted at face value. Equally amazing is the fact that the alleged motives of a second major beneficiary, business, are also uncritically accepted.

---

[40] Ibid.

[41] Rottenberg, "Minimum Wages in Puerto Rico," p. 337.

[42] Norman J. Ornstein, et al., *Vital Statistics of Congress 1982* (Washington, D.C.: American Enterprise Institute, 1982), p. 77.

[43] Ibid., p. 86.

Business Support for the Minimum Wage

At first glance it may seem strange that business would push for a higher minimum wage. On examination, however, we find that some businesses have the same motives for supporting the minimum wage as have labor unions. The economic self-interest of businesses that pay above the minimum wage dictates that they try to eliminate lower-cost rivals that pay below the minimum by forcing them to pay higher wage rates. That this is the case can be seen as far back as the time of the NRA codes.

Business interest groups definitely favored the minimum wage and price fixing of the NRA. In April 1935 the U.S. Chamber of Congress voted 1,495 to 419 in favor of continuation of the NRA.[44] After the Supreme Court declared the NRA unconstitutional, businesses collaborated with unions in lobbying for a substitute. The letter of C. R. Palmer, president of Cluett, Peabody and Co., to Rep. Arthur Healy (Mass.) on 25 March 1938 speaks for itself:

> When the present administration went into office, we wrote the Department of Labor, March 21, 1938, requesting that consideration be given to the possibility of establishing a minimum wage throughout the country. We realized that in many sections of the country wages were so low and hours so long that it made it impossible, or most difficult, to obtain business by companies that were paying good wages and working reasonable hours. So, when the N.R.A. provisions for wages and hours were adopted, we believed that objective had been obtained. Then when the N.R.A. collapsed, we were glad that many companies did not immediately increase their hours or reduce their wages. But now the situation is again difficult. . . . We had hopes that there would be a new bill controlling this situation within reasonable limits. . . . We are hoping that you agree with us, and that you will use your influence to see that something is done about it.[45]

Three months later, President Roosevelt signed the Fair Labor Standards Act into law.

When proposals to raise the minimum wage in the 1950s materialized, businesses were at the forefront of the lobbying effort. On 19

---

[44] Carpenter, *Competition and Collective Bargaining*, p. 759.

[45] *Congressional Record*, 28 March 1938, p. 1213.

July 1955, Representative Reuss from Milwaukee entered the following in the *Congressional Record:*

> I am proud of the fact that employers from my district have written me asking my support for the $1.25 minimum wage because they do not want sweatshop competition. Some of these letters from Milwaukee employers show a high level of social responsibility and business morality which I wish to quote for the Record:
>
> > Dear Congressman Reuss:
> >
> > This is to advise you that the Schmitt-Orlon Co. subscribes to the $1.25 minimum wage, as it will serve to bring closer the wage differentials between the low and high labor areas. . . .
> >
> > We, Wisconsin textile garment manufacturers, would then be placed on a better competitive basis.
> >
> > Yours very truly,
> > Arthur J. Schmitt,
> > Chairman of the Board.[46]

Businesses like the minimum wage for the same reason they like tariffs—to shut out lower-priced firms from free market competition. The minimum wage becomes a mechanism by which firms with high labor costs can force higher labor costs on their lower-cost competition. What makes it so seductive for businesses is that it can effectively close down their rivals, increase their incomes, and enable them to claim "social responsibility and business morality" in the process.

The Poverty Industry

The tragedy of the minimum wage is not only the 40–50 percent black teenage unemployment rates in our large cities, but that many drop out of the labor force altogether to become another poverty statistic or turn to criminal activity. If frustrated workers can end up as recipients of the generous welfare state, they may decide that receiving welfare is better than working productively. As Keith Leffler states:

> The alternative implicitly in mind for disemployed workers is zero income. However, this is unlikely to be the relevant alternative in the 1970's since public relief programs are available for nearly all categories of citizens laid off due to a higher minimum wage. Minimum wages may therefore be a technique of lowering the costs of establishing eligibility for the increasingly generous public welfare programs.[47]

---

[46] *Congressional Record,* 19 July 1955, p. 10984.

[47] Keith Leffler, "Minimum Wages, Welfare, and Wealth Transfers to the Poor," *Journal of Law and Economics* 21 (October 1978): 346.

Examining the ratio of average income from work at the minimum wage to average income from Aid to Families with Dependent Children (AFDC) welfare payments, Leffler found that "[o]n the average, AFDC payments alone replace over 80 percent of work income. In addition, work related expenses (excluding day care) averaged over 15 percent of earned income for those with minimum-wage income levels."[48]

Although increases in welfare payments tend to draw people away from work and onto the welfare rolls, it is not true that increases in the minimum wage necessarily draw people away from welfare.[49] Linda Leighton and Jacob Mincer make an analogous observation concerning the so-called inducement hypothesis, which describes the alleged incentive effects of the minimum wage on school enrollment: "It is clear now, given the evidence on labor force participation and on enrollment effects of minimum wages, that the inducement argument is not valid. Indeed, the logical conclusion is to the contrary: minimum wages induce welfare, not work."[50]

In this case, the disemployed workers who receive welfare are not the only vested interests; social workers and the entire "poverty industry" also benefit. If there are more poor because of the minimum wage, the demand for people to manage and take care of the poor also increases.[51] There is little incentive for social workers or poverty rights activists to advocate the repeal of the law that keeps them employed.

Politicians who "champion" a large poverty class are beneficiaries of the minimum wage as well. Since the law cuts off the first necessary condition for upward mobility (first-time employment), a permanent supporting constituency is preserved. A powerful triangle then emerges among the "poverty politician," social workers, and welfare recipients. Politicians who claim to represent the working poor can raise the minimum wage, throwing low-wage earners out of work. The modern welfare state can pick them up and provide them with greater benefits than what they could receive by working. Welfare recipients reward the politicians by voting for them, and the social workers and welfare bureaucrats find the demand for their services increased, entrenching their jobs. Each group in the political triangle of the minimum wage tends to benefit at the expense of the general taxpayer and the relatively low-skilled worker.

---

[48] Ibid., pp. 353–54.

[49] Ibid., p. 350.

[50] Linda Leighton and Jacob Mincer, "The Effects of Minimum Wages on Human Capital Formation," in Rottenberg, *Economics of Legal Minimum Wages,* p. 157.

[51] Walter Williams, "Economist Addresses Racial and Economic Fallacies," *Laissez-Faire Society,* 11 April 1984, p. 2.

IV. Beneficiaries and Victims of the Minimum Wage: A North-South Perspective

The beneficiaries and victims of the minimum wage are clearly visible when the historical attitudes of the Northern and Southern United States are compared. The North, tending to have higher wage rates due to earlier industrialization, would be the net beneficiary; the South, with lower wage rates, would be the major loser.

With the creation of the NRA, the United States had for the first time a national law for minimum wages. The minimum wage provisions of the NRA were not unified into a single wage for all classes of industrial labor; some were set differently.[52] Industries located in high-wage districts, in this case the North in general, desired an absence of wage differentials during the process of code formation.[53]

It has already been established that the minimum wage imposed by the NRA devastated the South and Puerto Rico. Benjamin Anderson summarizes the consequences of the NRA in the following observation:

> NRA created a great deal of unemployment in the South. Despite the differentials, the greatest percentage increase in wages, the greatest shortening of hours, and the greatest percentage increase in labor costs under NRA were in the South. In the lumber trade, for example, very little increase was made in western wages, while a very great increase was made in southern wages, even though they remained well below those of the West. . . . An interesting book by Charles Frederick Roos . . . gives significant figures in connection with this point and says: "In view of these data, it is not surprising that lumber business was diverted from the South to the Pacific Northwest."[54]

With the demise of the NRA, textile, apparel, lumber, and other industrial sectors began to experience vigorous competition from the South. This occurred because geographical wage differentials reappeared, reflecting underlying free market comparative advantages of doing business in the South. Consequently, Northern unions (such as the International Ladies Garment Workers Union) and industrial associations (such as the National Dress Manufacturers Association) desired the protection of Northern industry.[55] During the period before the enactment of the FLSA, both unions and businesses in the North had begun to express a fear of Southern competition.[56]

---

[52] Lyon, *National Recovery Administration*, p. 317.

[53] Ibid, p. 324.

[54] Anderson, *Economics and the Public Welfare*, p. 336.

In the needle trades, it became evident that the wage rates set by the NRA would not hold. Manufacturing associations attempted to re-create the NRA privately and failed completely, largely due to enforcement failures.[57] Such unions as the Amalgamated Clothing Workers of America attempted general organization drives with special emphasis on the South but also met with widespread failure.[58] Faced with these failures in attempting private cartelization to protect Northern from Southern industry, government cartelization became the logical alternative. This form of cartelization came in the form of the FLSA, as noted by Walter Boles: "There is evidence to show that some of the support behind the Act—a consequence of the interstate nature of markets—was based on the older industrial areas' fear of the growing industrialization of the South."[59]

The political-geographical background of the FLSA became sharply distinct. Northern senators and congressmen lined up to support the act, while representatives from the South opposed it. John F. Moloney observes that

> The elimination of so-called "unfair competition" and of regional wage differentials was also a factor of some importance. The lower wage structure of the South received its fair share of attention during the debates in Congress.[60]

Since businesses that pay above the minimum wage urge higher minimums on their competition, and since firms in the Northern United States tended to incur higher labor costs than their Southern counterparts, it is only logical that Northern firms would have pressed for passage of the FLSA. Paul Douglas states:

> So far as the employers were concerned, the northern textile industry was definitely in favor of the bill, and opposed to the granting of any regional differentials. It welcomed a national scale as a means of protecting themselves against southern competition with lower wages.[61]

---

[55] Carpenter, *Competition and Collective Bargaining*, pp. 812–13.

[56] *Congressional Record*, 28 March 1938, p. 1213.

[57] Carpenter, *Competition and Collective Bargaining*, p. 815.

[58] Ibid.

[59] Walter Boles, "Some Aspects of the Fair Labor Standards Act," *Southern Economic Journal* 6 (April 1940): 498–99.

[60] John F. Moloney, "Some Effects of the Fair Labor Standards Act Upon Southern Industry," *Southern Economic Journal* 8 (July 1942): 15.

[61] Douglas, "Fair Labor Standards Act," p. 501.

According to Senator Walsh (Mass.), the new minimum wage was to provide protection for Northern industry:

> Industries in New England that now operate on a forty hour week . . . and pay reasonable minimum wages will not continue to be subject to competition in the trade markets with goods produced by industries that can undersell New England producers due to the fact that they are working their employees longer hours and paying lower wages."[62]

A telegram sent to Representative Healey from Governor Charles F. Hurley of Massachusetts demonstrates the fact that the FLSA was favored in the North because of the protection it afforded. The telegram did not hide the fact that the law was in the interests of this particular Northern state.

> We here in Massachusetts are deeply interested in having the pending wage and hour bill adopted by Congress during this session. As a Massachusetts Congressman I am sure you realize how important it is for such Federal legislation to be adopted and for Massachusetts to have equal competition with other sections of the country, thus affording labor and industry of Massachusetts some degree of assurance that our present industries will not move out of the State.[63]

And on 6 April 1938, Rep. John F. Dockweiler (Calif.) openly admitted that he could not think of a single trade or industry in his state where the prevailing minimum wage was not considerably higher than the 30 cents per hour proposed in the FLSA.[64] Of course, he then went on to throw his unbending support for its passage. On 4 May 1938, Rep. Robert Allen (Penn.) introduced in the *Congressional Record* a radio address of Representative Healey from the previous day. Healey's comments again show his concern for Massachusetts:

> Sweatshops and low-wage areas have grown like mushrooms in all sections of the country. . . . They have . . . undermined decent industry by ruinous competition, snatched away markets, lured factories out of high-standard areas. . . .
>
> . . . Massachusettes, long preeminent in the manufacture of textiles, has seen its commanding position swept away by the corrosive competition of sweated industries. Between 1923 and 1933, the New England textile industry lost nearly 120,000 jobs, mostly from Massachusettes. One hundred and twenty thousand jobs were taken out of that industrial region, principally because of low wages elsewhere. . . .[65]

---

[62] Ibid.

[63] *Congressional Record,* 3 May 1938, p. 1800.

[64] *Congressional Record,* 6 April 1938, p. 1353.

[65] *Congressional Record,* 4 May 1938, pp. 1840–41.

We find congressmen from Indiana making similar statements:

> There are in the state of Georgia, canning factories working ... women 10 hours a day for [$]4.50 a week. Can the canning factories of Indiana and Connecticut or New York continue to exist and meet such competitive labor costs?[66]

A common explanation for the passage of the FLSA—the protection of the helpless working poor—is blatantly deficient in that it completely ignores the geopolitical struggle that pervaded the issues.[67] The driving force behind the establishment of the minimum wage was Northern unions and businesses. They lobbied hard for the law and were successful.[68] To the delight of such unions as the International Ladies Garment Workers Union and the Amalgamated Clothing Workers of America, the law completely destroyed the Puerto Rican needle trades. Owners of Northern textile firms who watched the government drive their competition out of business expressed little regret at that outcome.

Henry Simons demonstrated keen perception of the entire minimum wage issue some years later when he declared:

> Southern workers may be intrigued by the wage expectations held out by organizers from northern unions and by the Fair Labor Standards Act. They may in a few cases get such wages, but it will be only in spite of the intentions of the northern unions and Massachusetts senators.[69]

The objections raised by Southern congressmen that the minimum wage was detrimental to the economy of the South were always conveniently pushed aside by politicians representing the North. In January 1954, Sen. John F. Kennedy published an article entitled "New England and the South: The Struggle for Industry" in the *Atlantic Monthly.* In that article Senator Kennedy made clear his belief that raising the minimum wage would help stem the migration of industry from Massachusetts to Southern states such as Mississippi:

> But the final reason for migration, with which I am particularly concerned, is the cost differential resulting from practices or conditions permitted or provided by Federal law which are unfair or substandard by any criterion. Massachusetts manufacturing indus-

---

[66] Grossman, "Fair Labor Standards Act," p. 27.

[67] Boles, "Some Aspects of the Fair Labor Standards Act," p. 499.

[68] S. Charles Maurice and Margaret Jane Hobson, *Minimum Wage Laws: Who Benefits, Who Loses?* (College Station, Tex.: Center for Education and Research in Free Enterprise, Texas A & M University, 1983), p. 18.

[69] Henry Simons, *Economic Policy for a Free Society* (Chicago: University of Chicago Press, 1948), p. 135.

> tries in May of 1953 paid an average hourly wage of $1.64; but because the Federal minimum is only 75 cents an hour, many industries migrating to the rural communities of Mississippi pay workers only that less-than-subsistence wage, and those employees under "learners permits" even less. Practically all New England woolen textile mills pay a wage of at least $1.20 an hour; but . . . the New England Mills must bid for government contracts against southern mills paying only $1.05 an hour.[70]

The 1955 debates over increases in the minimum wage were the sharpest North-South confrontation on this issue. With reference to Puerto Rico, President David Dubinsky of the International Ladies Garment Workers Union urged a congressional subcommittee to raise the minimum wage. The reason he gave was that the Puerto Rican brassiere industry was growing and that Northern knitting mills had begun to move to the island.[71]

In the 19 July 1955 House debate Representative Mack (Wash.) reveals why he supported an increase:

> Mr. Chairman, in the three Pacific coast states of Washington, Oregon, and California about 600,000 men are employed in forest-product industries. . . . The lumber, plywood, shingle, veneer, doors, and furniture which these workers produce must be sold in competition with similar products made by workers in southern states. . . . The southern manufacturers thereby have a competitive advantage over our west coast mills, where wages are higher. . . . It is not fair that western producers of lumber, plywood, furniture, and other forest products who pay an average of $1.80 an hour must compete with the southern lumber, plywood and furniture manufacturers who pay an average wage of only about 86 cents an hour. This unjust differential can be remedied by requiring that southern manufacturers pay at least a minimum wage of $1 an hour.[72]

In the same House debate, Representative Nelson (Maine) summarized his feelings about the minimum wage in these words: "Today the statutory minimum wage has a far broader purpose than that. . . . In brief, it is a salutary and excellent device to prevent competition between states and sections of this country."[73] He also conveyed the findings of a committee appointed by the Conference of New England Governors. They found that the major explanation of New England's decline in textiles was the large wage differential between New England and the South.[74]

---

[70] John F. Kennedy, "New England and the South: The Struggle for Industry," *The Atlantic,* January 1954, p. 33.

[71] Rottenberg, "Minimum Wage Laws in Puerto Rico," p. 338.

[72] *Congressional Record,* 19 July 1955, p. 10978.

[73] Ibid., p. 10961.

In 1960, Representative Lindsay (N.Y.) reminded his colleagues that he was supporting an increase in the minimum wage to protect the apparel industry in his district.[75] In 1966, Representative Resnick (N.Y.) favored increases in the minimum wage to protect farmers in the Northeast who paid workers $1.25 to $1.75 per hour from farmers in Mississippi who paid only $3.00 per day.[76]

Marshall Colberg, who views the minimum wage as a basic tool of the North to stifle Southern competition, recently tested North-South records on key votes. He looked first at the voting on the 1961 Monroney Amendment, which would have limited the coverage of the minimum wage for retail, gasoline station, laundry, and construction workers to firms established in more than one state. Colberg found that senators from Southern states voted 75 percent in favor of the amendment, while Northern senators voted 75 percent against it.[77]

Surprisingly, however, Colberg also found that Southern senators voted by a large majority in favor of a 1966 bill extending minimum wage and overtime provisions to a larger number of workers in the District of Columbia. Indeed, "even such strong opponents of the federal minimum wage as Spessard Holland of Florida, Everett Dirksen of Illinois, and Sam Ervin of North Carolina voted yea. Evidently, voting on the minimum wage is not strictly a matter of principle."[78] In what has been called the "Colberg hypothesis," Colberg determined the anomaly was only apparent: legislators from more industrialized areas—including relatively more industrialized areas in the South—will always support the minimum wage as a device to reduce business migration from their districts to districts with lower wages. This hypothesis has found support in a test conducted by Jonathan Silberman and Gary Durden.[79]

---

[74] Ibid. Representative Gwinn (N.Y.) had some interesting comments to his colleagues that are still relevant today. He asked (ibid., p. 10955):

> Where did Congress ever get its power except to seize that power, to say to a working man in America, "You shall not work for less than 75 or 90 cents an hour." That is the way this statute reads, in effect. It is not just that the employer must not pay less than 90; the employee must not work for less than 90.

[75] *Congressional Record*, 30 June 1960, p. 15212.

[76] Marshall Colberg, "Minimum Wages and the Distribution of Economic Activity," in *Economics of Legal Minimum Wages*, edited by S. Rottenberg, p. 249.

[77] Ibid., p. 251.

[78] Ibid., p. 253.

[79] Jonathan Silberman and Gary C. Durden, "Determining Legislative Preferences on the Minimum Wage: An Economic Approach," *Journal of Political Economy* 84 (April 1976): 327.

V. The Prospects for Change

Before one can hope for the abolition of the minimum wage, the influence of the well-organized, special-interest groups that receive the bulk of the benefits from it must be reduced. There is, however, little chance that such groups as Northern unions and businesses, or interest groups in general, could be barred from the political process in the near future. Other options for change would include (1) payoffs to the vested interests, (2) two-tiered minimum wages, (3) inflation, and (4) general economic education of voters and victims.

Although paying off vested interests is theoretically possible, it would not seem very promising for two resons. First, in order to compensate Northern unions and firms for their lost rents, taxes would have to be increased. This would merely shuffle the burden of the law, shifting it from teenagers to the general taxpayer. Second, it may be politically impossible to make direct transfer payments to these groups because the visibility of such payments could put the entire rent-seeking scheme in jeopardy. Since the societal costs of the minimum wage law could no longer be hidden, the payoff scheme would have to include a high risk premium to compensate for the probability of abolition by angry taxpayers.

Two-tiered minimum wage rates have been recently proposed by President Reagan, who asked Congress for a sub-minimum wage of $2.50 per hour for teenagers for work done in the summer months.[80] In spite of evidence that such a minor reduction would result in the additional employment of 400,000–600,000 youths,[81] the sub-minimum wage is legislatively dead. In 1977, Congress also turned a deaf ear to pleas from private youth-employment centers for a lower minimum wage for younger workers.[82] The intellectual tide of opinion is moving toward favoring a two-tiered minimum wage, but it is not clear whether this movement will be translated into actual labor policy.

Inflation is sometimes considered an escape from the current minimum wage dilemma because it erodes a fixed minimum wage, eliminating its effectiveness. But this "method" poses serious problems. By using inflation to reduce minimum wage–created unemployment, greater unemployment could be created if in the process of monetary expansion relative prices are distorted in the structure of production, thus causing malinvestment.[83] That is, inflation could initiate a trade cycle, bringing with it an ensuing depression.

---

[80] Maurice and Hobson, *Minimum Wage Laws*, p. 18.

[81] Ibid., p. 33.

[82] *Congressional Record,* 6 October 1977, p. 32704.

Educating voters about the minimum wage may be the only lasting solution to the problem. The large number of victims must be made aware of the links between the minimum wage law and their plight. But this can be very difficult when politicians use the law to their political advantage and intentionally cloud the issue with emotional rhetoric. In trying to convince Americans of the necessity of the minimum wage provision in the FLSA, President Roosevelt used tactics like the following: "Do not let any calamity-howling executive with an income of $1,000 a day . . . tell you . . . that a wage of $11 a week is going to have a disastrous effect on all American industry."[84] Politicians have been extremely effective in making the minimum wage popular with well-intentioned voters.[85] Economists, on the other hand, have much work to do in the slow process of giving the average voter the economic education he needs to understand—and then abolish—the minimum wage.

VI. Conclusion

George Stigler may have startled some economists in 1946 when he claimed that minimum wage laws create unemployment and make people who had been receiving less than the minimum poorer.[86] Fifty years of experience with the law has proven Stigler correct, leaving very few defenders in the economics profession.[87]

But economists have had little success in criticizing this very destructive law. Simon Rottenberg demonstrated the government's disregard for what most economists have to say about this issue in his investigation of the Minimum Wage Study Commission created by Congress in 1977. He noted the numerous studies presented to the commission that without exception found that the law had a negative impact on employment and intensified the poverty of low-income earners. The commission spent over $17 million to conduct the investigation and on the basis of the evidence should have eliminated the law. What was the outcome? The commission voted to *increase the minimum wage by indexing and expanding coverage.* As dissenting commissioner S. Warne Robinson commented about the investigation:

---

[83] See Friedrich Hayek, *Unemployment and Monetary Policy: Government as Generator of the Business Cycle* (Washington, D.C.: Cato Institute, 1979).

[84] Grossman, "Fair Labor Standards Act," p. 22.

[85] William Keech, "More on the Vote Winning and Vote Losing Qualities of Minimum Wage Laws," *Public Choice* 29 (Spring 1977): 134.

[86] George Stigler, "The Economics of Minimum Wage Legislation," *American Economic Review* 36 (June 1946): 358–65.

[87] Although there are a few supporters left such as John K. Galbraith, many "liberal" economists such as Paul Samuelson and James Tobin have recently come out against the minimum wage. See Emerson Schmidt, *Union Power and the Public Interest* (Los Angeles: Nash, 1973).

> The evidence is now in, and the findings of dozens of major economic studies show that the damage done by the minimum wage has been far more severe than even the critics of forty years ago predicted. Indeed, the evidence against the minimum wage is so overwhelming that the only way the Commission's majority was able to recommend it be retained was to ask us not to base any decisions on the facts.[88]

It cannot be that our elected representatives in Congress are just misinformed with respect to the minimum wage law. To the contrary, the *Congressional Record* demonstrates that they fully understand the law's effects and how the utilization of those effects can ensure reelection. Economists would do well to realize that governments have little interest in the truth when its implementation would contradict self-serving government policies. Rather than attempting to bring government the "facts," economists should educate the public. This is the only solution to the malaise created when people uncritically accept such governmental edicts as the minimum wage.

---

[88] Simon Rottenberg, "National Commissions: Preaching in the Garb of Analysis," *Policy Review* no. 23 (Winter 1983): 139.

IN DEFENSE OF THE INVISIBLE HAND

Norman P. Barry

Introduction

Despite the current intellectual popularity of free market philosophies and the increasing doubt cast on the idea that the political power of the state can improve upon the spontaneous workings of a decentralized, private property exchange system in the production of desirable social outcomes, criticism of the Invisible Hand hypothesis continues to flourish. Moreover, such criticism does not emanate solely from the members of some socialist ancien régime anxious to preserve an intellectually spent economic dynasty but from some of the most original and sophisticated economic and social theorists currently writing. Two outstanding contributions to this critique of classical economic liberalism are those of Frank Hahn and Amartya K. Sen, whose Fred Hirsch Memorial Lectures have attracted much attention.[1] What is of interest to the Invisible Hand theory in these articles, especially that of Hahn, is not that its critics wish to amputate the limb and replace it with some allegedly superior steering device, but that they seek only to expose infirmities in its joints and muscles and indicate some unsureness of its grasp. Nevertheless, the hand of friendship proffered by these writers seems limp enough to provoke some defensive reflections on behalf of spontaneous market forces.

The criticisms of Invisible Hand processes cover a wide range of economic and social issues. At the level of pure theory, there is the question of whether rational economic agents, driven only by self-interest, will produce an economic "equilibrium" and, further, whether

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Professor of Politics at the University of Buckingham.

[1] See Frank H. Hahn, "Reflections on the Invisible Hand," *Lloyds Bank Review*, April 1982, pp. 1–21; and Amartya K. Sen, "The Profit Motive," *Lloyds Bank Review*, January 1983, pp. 1–20. See also Frank Hahn, "General Equilibrium Theory," in *The Crisis in Economic Theory*, ed. D. Bell and I. Kristol (New York: Basic Books, 1981), pp. 123–38.

"equilibrium" is a useful description of economic reality. With regard to welfare economics, the critics argue that the optima described by the theorems of market equilibrium have no necessary connection with economic justice, since the optima are consistent with any given distributions of property holdings. Sen, especially, challenges the utility of the market in relation to employment and welfare policy in industrialized countries and seriously questions its success in promoting prosperity in developing areas. Generally, both Hahn and Sen express skepticism about the possibility that economic agents can actually advance their personal interests by obeying the egoistic injunction of the laissez-faire theory. Sen at least implies that such a postulate does not capture a desirable or adequate ethical goal. This paper examines these criticisms of the Invisible Hand theory in light of Hayek's insights concerning the market process, and in light of the recent developments in public choice theory and property rights theory.

The Invisible Hand as a Competitive Market Process

The defense of the Invisible Hand process must begin at the theoretical level with an accurate and coherent account of what is being claimed in the economic philosophies of those who utilize it. For it can be argued, with ample justification, that much of the initial plausibility of the criticism of the Invisible Hand theorem derives from a systematic misrepresentation of the spontaneous market order that it describes. In fact, the description of a competitive market system given in the orthodox texts of neoclassical economics bares little resemblance to the theory of competition advanced by Adam Smith and his successors. The differences center on the conceptions of human action that underlie rival models of economic competition.

The difficulty of understanding the significance of the Invisible Hand theorem is that, in economics at least, it has become inextricably bound with the notion of equilibrium: the search for the presence of an Invisible Hand process has been misdirected toward the search for the existence of equilibrium in a decentralized market economy. To anticipate what is to follow, the critics of the market system claim that the absence of a stable equilibrium in real economies entails the refutation of the Invisible Hand theorem. I maintain that this is false.

The Shortcomings of Hahnian Equilibrium

Although Hahn is a distinguished exponent of general equilibrium economics, it is not always clear what exactly he means by the concept or how it is to be used in economic analysis. However, the policy implications he draws from equilibrium theory are fairly clear.

In his Fred Hirsch Lecture, Hahn appears to be describing the "pure theory" of the Invisible Hand as equilibrium theory. An economic system is postulated in terms of a perfect coordination of economic actions in which the price structure reveals no possibility of further changes. According to Hahn:

> The economic environment of any one person is fully specified once the prices of all tradeable objects are given. These prices are the terms at which one good can be exchanged for another and it is a basic assumption that all individuals can trade to any extent they wish at these prices. One notices that the economic information is conveyed very economically—the individual knows everything that he needs to know once he knows prices.[2]

From this it follows that all individuals are price-takers (no one person can influence price), they are endowed with perfect knowledge of all possible states of affairs, and markets are "complete," that is, there are terms of exchange that enable individuals to make any trades they wish. Individuals then react in the manner of automatons to their environment, and their actions can be predicted by an external observer. These actions are, in Hahn's own words, "machine-like responses of agents to prices."[3] The argument therefore for the operation of the Invisible Hand is reduced to the abstract *mathematical* proof of the proposition that in a world of decentralized trading the actions of transactors will "mesh"; there will be no discoordination to be corrected by further trading.

A more adequate statement of Hahn's position on equilibrium can be found in his earlier essays,[4] where two sorts of equilibrium are differentiated: the timeless static world described in the original Arrow-Debreu[5] model and a more dynamic version in which the sequential movement of an economy is explored. Although no different policy conclusions follow from the two representations, a fair treatment of Hahn's views requires that they be examined.

The latter is much more interesting and, as S. C. Littlechild has pointed out, has much in common with the Hayekian[6] version of equilibrium. It represents a significant departure from the original Arrow-Debreu model, for instead of using equilibrium theory to represent or "objectify" a static and unchanging picture of perfect coordination, it addresses the question of how individuals, through a learning process, adjust their plans through time so as to bring about that equilibrium state. This "subjectivist" view of economics sees the task of equilibrium theory as one of incorporating an explanation of individual action in a less rarefied world than that of comparative statics. Hahn puts the matter accurately when he says that this "requires that information processes and costs, and also expectations and uncertainty be essentially included in the equilibrium notion. This is what the Arrow-Debreu construction does not do."[7]

---

[2] Hahn, "Reflections on the Invisible Hand," p. 2.

[3] Ibid., p. 6.

[4] In particular, see Hahn, "The Winter of Our Discontent," *Economica* 40 (August 1973): 322–30; and his inaugural lecture at Cambridge University, *On the Notion of Equilibrium in Economics* (Cambridge: Cambridge University Press, 1973).

[5] K. J. Arrow and G. Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica* 22 (July 1954): 265–90.

[6] F. A. Hayek, "Economics and Knowledge," in *Individualism and Economic Order* (London: Routledge and Kegan Paul, 1948).

In Hahnian equilibrium, which sees economic action as sequential action, each individual has a theory of how the economy will develop over time, and his theory will be modified (through a learning process) in accordance with his perception of changes in the data that confront him. If every change is perfectly foreseen then each individual is in equilibrium: An economy is in equilibrium when all the individual theories or plans are mutually compatible. Thus, Hahn's notion of equilibrium is similar to that used by Hayek in the *Pure Theory of Capital.* Hahnian equilibrium is undoubtedly an improvement on the Arrow-Debreu model, which, by objectifying a once-and-for-all state of affairs, is unable to accommodate change. But it is not at all clear that Hahn's modified version improves our understanding of how real markets work or contributes to an explication of the Invisible Hand theorem.

The crucial point is that Hayek sees the *limitations* of equilibrium theory, as Mises did before him (although the Misesian construction of an "evenly rotating economy" is less useful since it does not accommodate change). A theory of equilibrium is a deliberate abstraction from reality, a limiting case, which enables us to understand how and in what ways the behavior of real economies departs from the theoretical model. Thus, far from extending and refining the basic postulates of equilibrium theory so that they provide a truer picture of reality, Hayek goes on to take account of other economic mechanisms.[8] Notable here, of course, is entrepreneurship (completely absent from any equilibrium model), which explains how an economy tends to move toward that state of perfect coordination described by abstract theory.

---

[7] Hahn, *On the Notion of Equilibrium in Economics*, p. 16. See also, S. C. Littlechild, "Equilibrium and Market Process," in *Method, Process, and Austrian Economics*, ed. Israel Kirzner (Lexington, Mass.: D.C. Heath, 1982).

[8] See especially, "The Meaning of Competition," in *Individualism and Economic Order.* The most sophisticated contemporary competitive theory is Israel Kirzner's *Competition and Entrepreneurship* (Chicago: University of Chicago Press, 1973).

But Hahn believes, mistakenly, that further progress in economic science can only come from the continual refinement of the general equilibrium model. While accepting (emphatically) that this does not reflect reality, he claims that "The student of general equilibrium believes that he has a starting point from which it is possible to advance towards a *descriptive* theory" (italics added).[9] Thus Hahn is not interested in economic processes, *genuine* competition, entrepreneurship, or innovation, all of which characterize real economies, but in routine, regularized, and repetitive behavior that can be encapsulated in equations. As Alan Coddington points out,[10] this is entirely because Hahn is wedded to the formal deductive method in economic theory. But the real work of the Invisible Hand is done in those coordinating market processes that cannot be captured by axiomatic reasoning. Hahn shows some recognition of this when he concedes that "general equilibrium is strong on equilibrium and very weak on how it comes about,"[11] but he never takes up the market process theory that is hinted at here.

What is worse, as again Coddington points out, Hahn uses theory in a systematically misleading manner.[12] Although he occasionally concedes that general equilibrium is a self-contained (mathematical) exercise that does not picture reality, he uses it to "falsify" the major claim of the Invisible Hand theory: that an unhampered, decentralized market economy will tend to produce a Pareto-optimum. He does this by showing that, in comparison to the perfect coordination of plans described by pure theory, existing markets display many "inefficiencies" that could in theory be eliminated. The argument, however, is misleading because general equilibrium theory is not a causal explanation of how markets work but rather a self-contained logical exercise that deliberately excludes coordinating processes.

Implicit also in Hahn's analysis is a skillfully concealed piece of normative economics. What he wants to show is that in a variety of areas the (alleged) failures of the market—in terms of the standards set by abstract theory—can be corrected by government and administrative action, presumably informed by the equations of general equilibrium theory. What he does not consider is that although existing markets may not provide "perfect" solutions to the problems of, say, exhaustible resources, externalities, public goods, and so on, this still leaves "wide open the possibility that the imperfectly-functioning markets perform better than any of the available (administrative) alternatives."[13]

---

[9] Hahn, "Winter of Our Discontent," p. 324.

[10] Alan Coddington, "The Rationale of General Equilibrium Theory," *Economic Inquiry* 13 (December 1975): 547.

[11] Hahn, "Winter of Our Discontent."

[12] Coddington, "Rationale of General Equilibrium Theory," pp. 541–43.

Thus even though the full description of Hahnian equilibrium includes some subjectivist elements, it no more captures the essence of the Invisible Hand theorem than does the static model he is so anxious to refine. He has not produced an account of its movement at all but engineered a sophisticated sleight of hand. His sequential equilibrium is still a description of the outcome of a market process and not an account of the process itself. An economic "end-state" has been postulated after the Invisible Hand has done its work, but what is not revealed is how this work is done.

The Invisible Hand Process

The notion of the Invisible Hand must be seen as a metaphor that illuminates a continuing process of exchange and competition between individuals which brings about a coordination of plans and purposes. It must not be seen as a picture of an end-state of perfect equilibrium in which all plans have already meshed, since that implies the cessation of human action. The Invisible Hand image refers to an unending process of change and adjustment and not to a perfectly harmonious end-state in which incentives to change have been removed. It is because we cannot know in advance of a market process the details of the price structure of an economy in perfect equilibrium that the description of a market economy in those terms is quite misleading: It is merely a theoretical abstraction that has no necessary connection with real markets.

Of all the differences that mark off abstract equilibrium models from existing markets it is the omnipresence of ignorance and uncertainty that is most germane to real markets. Because market transactors are necessarily ignorant of most of the facts of the economic universe, all of their actions are speculative; they make guesses about a necessarily unknowable future. From this it follows that they cannot be passive price-takers responding automatically to a given set of data, but are active makers of various futures. Economic activity consists of continually exploiting price differences that exist in a necessarily imperfect world. It is this, of course, that constitutes profit, a phenomenon which is absent from an equilibrium world. Yet if there were no possibility of profit there would be nothing to drive the system toward equilibrium. In Hayek's instructive phrase, the market is a "discovery procedure" by which transactors adjust to ever-changing circumstances, rather than an "allocative device" by which means are somehow mechanistically directed to the production of given ends.

---

[13] Ibid., p. 554.

We can now see how genuine market competition differs from the ersatz version described in "perfectly competitive equilibrium" models. As Frank Knight said, "in perfect competition there is no competition." If there are no mistaken prices to correct, as is the case with perfect competition, there is no opportunity for the intense rivalry that characterizes economic action to manifest itself and no need for entrepreneurship to play its coordinating role.[14]

In economics the point of the Invisible Hand theorem is to show how there can be order without a designing mind and without anyone intending specifically to produce such an order. Hayek's famous observation that the social sciences should be concerned with the investigation of phenomena that are "the result of human action, but not of human design"[15] shows there are "natural" processes at work that, if left undisturbed, will produce an order infinitely more complex than that which emanates from deliberate human will. This is because no one mind can have access to that dispersed knowledge which is a feature of a natural system: To think this is possible is, according to Hayek, to be a victim of the "synoptic delusion." As Hayek has been at pains to point out, a self-correcting economic system is not the only example of a natural social process; legal systems and languages, for example, display similar properties.

Sen's Objection to Spontaneous Order

Sen thinks that this is rather an "unprofound" thought.[16] He supports this contention by giving a trivial example of an action—crossing the street—from which certain results occurred that were not specifically designed; for example, crossing the street led to a passing car being delayed. This is quite disingenuous, for the fascinating thing about Invisible Hand theories is that they produce surprising and *untrivial* results. The typical modern intellectual is a victim of the synoptic delusion: It is inconceivable to him that a market can coordinate in the absence of a central human agency, or that judges, in a case-by-case manner, accidentally generate a more predictable legal order than that produced by a legislature. Yet it is these processes that are orderly and the human will that is capricious.

---

[14] F. A. Hayek, "Competition as a Discovery Procedure," in *New Studies in Philosophy, Politics, Economics and the History of Ideas* (London: Routledge and Kegan Paul, 1978), pp. 179–90.

[15] See Hayek, *Studies in Philosophy, Politics and Economics* (London: Routledge and Kegan Paul, 1967), pp. 96–105.

[16] Sen, "Profit Motive," p. 3.

The most that Sen concedes to the market is that it is appropriate only for those matters over which people's interests converge, but that it is quite irrelevant for those areas where there is a conflict of interests. In one sense, Sen is uttering a tautology: market relationships *are* convergent ones and, where there are irreconcilable conflicts, trading is impossible. The interesting point about the familiar institutions of market society, however, is that they enable individuals to find out those areas where cooperation and gains from trade are possible. We cannot know what coordination can take place until we allow people to exercise their "natural propensity to truck, barter and exchange." The danger of overemphasizing the conflictual side of human relationships is that it licenses "politics" to dominate men's lives, and politics has an almost irreversible tendency to conceal the opportunities for agreement among people.

Normative Implications of Conventional Invisible Hand Theory

We are now in a better position to scrutinize more critically the normative implication, noted above, that lurks in the complexity of the conventional version of the Invisible Hand theory. The obvious imperfections, such as monopoly power, of real-world markets are held up as departures from some imaginary social optimum crying out for governmental correction. But if no social optimum exists independently of the actions of economic transactors, how can there be an infallible touchstone that measures so exquisitely all of our economic variables? The familiar market imperfections, since they represent opportunities for speculative gain, are themselves being constantly corrected by enterprising individuals through the market process. If there is nothing in economic behavior apart from the actions of individuals, from where do we derive an instrument for callibrating those actions?

It follows then that there are two problems for the general equilibrium versions of the Invisible Hand theorem. First, there is the epistemological argument that the nature of free economic activity, which is essentially unpredictable, precludes any observer having the *knowledge* needed to make any statement about a social optimum meaningful. Second, even if the idea of such an optimum could be made operational, how would we guarantee that government action would produce it more effectively than private agents? Why should political officials be more informed about profitable opportunities than market traders who risk their own resources? Furthermore, as the public choice school of political economy has shown, we cannot be at all confident that political officials will maximize such social optima in the absence of strict constitutional rules. In fact, the reverse is likely to be the case, that is, officials will maximize their private interests.

The market is a process that operates through time; it is not characterized by the instantaneous adjustment of carefully programmed automatons. If the Invisible Hand's operation were that described in the general equilibrium theory, there would be no need for money, the firm, entrepreneurship, and all the other economic categories that exist only because of the ineradicable uncertainty that pervades economic life in the real world. The interesting question concerns not the existence of perfect coordination in abstract equilibrium, but the nature of the coordinating process that the Invisible Hand generates.

Here there would appear to be some disagreement between the various anti-equilibrium schools of political economy. The Austrian tradition, as exemplified by Mises and Hayek, holds that there is a tendency to equilibrium: This does not consist of an instantaneous price and quantity adjustment, but of a learning process in which individual plans are coordinated. However, a more extreme subjectivist position, associated mainly with the work of G. L. S. Shackle,[17] holds that because economic life consists entirely of thoughts and expectations about the future, which may turn out to be wrong, one cannot even posit a tendency to equilibrium. The world is "kaleidic" rather than smoothly coordinating, and the entrepreneur is a creative innovator rather than an agent that merely brings about an alignment of dispersed information.

The failure of the Invisible Hand derives not from considerations advanced by Hahn and Sen, but from a subtle interpretation of the very process that their models conceal. We cannot here go into the intricacies of Shackle's position, but experience strongly indicates that the disruptions to the coordinating process that do occur come from government intervention rather than from some endogenous features of the exchange system. Throughout the 20th century, government mismanagement of money has had catastrophic discoordinating effects, and the plethora of welfare and trade union legislation has systematically rendered large parts of the labor force virtually immobile. In other words, the Invisible Hand is not naturally arthritic: It has artificially imposed impediments that confine its movements.

---

[17] See his *Epistemics and Economics* (London: Cambridge University Press, 1972).

Critics of the Invisible Hand theorem sometimes concede that under specified conditions, a decentralized exchange process will generate a Pareto-optimum, a state of affairs in which no reallocation of resources is possible without making (at least) one person worse off.[18] But they are quick to point out the limitations of this notoriously weak principle of welfare economics. It is, of course, nice to know that unplanned markets can generate Pareto-improvements, but the potential beneficence of this principle is severely qualified by the fact that it is silent on the morality, or otherwise, of the distribution of property rights from which trading is to start. Of course, there are as many Pareto-optimal outcomes as there are possible property distributions. As Sen rightly points out: "A state in which some people are suffering from acute deprivation while others are tasting the good life can still be Pareto-optimal if the poor cannot be made better off without cutting into the pleasures of the rich."[19]

This is a much labored objection to free market economics, but its significance is vastly overstated. All it says is that if the Invisible Hand mechanism is to function as something more than a scientific explanation of economic order, it must produce separate principles that establish the legitimacy of the original property holdings of market traders. As a matter of logic, an exchange process must begin with objects that were acquired by means other than exchange. This, however, is not a special problem of the Invisible Hand explanation but a general problem of justification in politics and ethics.

Invisible Hand theorists become political liberals when they produce logically independent arguments for the justification of individual acquisition of previously unowned resources, of inheritance, gifts, and other forms of entitlement. Some of these arguments may not be persuasive, but can it really be maintained that socialist arguments for just acquisition are morally superior? Most of these, in fact, maintain the moral fiction that the "community" can somehow own valued things, a proposition that, in practice, has subverted not only the morality of individual freedom but also those individualistic mechanisms that power the exchange system.

Irrespective of these more substantive liberal considerations, one important indirect political implication of the classical liberal version of the Invisible Hand theorem is scarcely considered at all. This is the view that we cannot simply choose *any* distribution of property rights, since the choice is likely to have unintended consequences that are impossible to foresee.[20] This argument derives from the antirationalist version of the Invisible Hand theory and presses home the case that there are self-correcting mechanisms at work—not merely in the market economy but in the legal and social system generally—that produce outcomes which an unaided reason is powerless to improve upon. A good example is the evolving common law system, which has historically coordinated an individual's actions much better than the imposed statutes of rationalist planners. It is because of our ignorance that we cannot plan the future and, therefore, in a utilitarian sense, it is advisable to accept inherited institutions. This would dispose us then toward accepting a received system of property rights, which in most cases has developed accidentally, rather than planning them anew according to some arbitrarily imposed abstract principle.

---

[18] Hahn, "Reflections on the Invisible Hand," p. 4.

[19] Sen, "Profit Motive," p. 6.

Although the extension of the Invisible Hand theory appears to involve a descent into an uncritical traditionalism, which is too conservative for radical laissez-faire economists, its importance in the debate about the institutional framework within which market exchanges take place is frequently overlooked. Even those who favor free exchange and production for the satisfaction of *individual* wants (such as the "market socialist" thinkers of the 1930s) blithely presuppose that property rights can be shuffled and reshuffled endlessly so as to produce some socialist Pareto-optimal end-state without there being any adverse effects on the institutional framework of economic society.

Alleged Deficiencies of the Invisible Hand

Much of the contemporary criticism of the Invisible Hand theory is directed less to theoretical concerns and more to the alleged real-world deficiencies—economic, political, and ethical—of a system based on the spontaneous interaction of basically selfish economic agents. The argument here concerns the need for state action to correct an unhampered market that would otherwise periodically produce mass unemployment, pockets of poverty amid affluence, "externalities" in the form of pollution and damage to the environment, and fail to generate economic growth in Third World countries. In all this there is scarcely any recognition of the obvious fact that no other economic system has remotely approached capitalism in its productivity and ability to satisfy consumer wants. Centrally planned systems not only depend on capitalist systems for vital technological know-how but also, and more important, for the whole range of pricing signals provided by the international market system. In the absence of the latter, the productive process in a planned system would lack any kind of rationale.

---

[20] See Hayek, *The Constitution of Liberty* (London: Routledge and Kegan Paul, 1960), pp. 124–30.

In light of this, it is incredible for Sen to claim that, in the market economies of the West, people do not go begging for food because of the "social security that the state has offered."[21] It is surely obvious that the high welfare payments in capitalist economies are paid out of a surplus created by an economic system driven by the profit motive. Indeed, the social security payments in capitalist economies in many cases exceed the incomes of employed workers in plannned economies. Again, it is neither naive nor callous to suggest that much of the unemployment that characterizes market economies is a result of those very same welfare payments that cramp the movements of the Invisible Hand. People who choose subsidized leisure over paid employment are responding rationally to the signals of a distorted market. It is intellectual duplicity to argue that economic inefficiencies caused by interventionist inhibitions in the working of the Invisible Hand constitute evidence of its failings.

Sen[22] pays some attention to the version of market theory that stresses the inviolability of the property rights of the participants to the exclusion of any consequentialist considerations that would normally be thought relevant to an evaluation of the outcome of the exercise of those rights. Thus he concludes that actual instances of mass starvation in developing countries can be attributed to the property rights structure of market society. Here, the Invisible Hand is said to fail in a welfare sense because it does not distribute food, in conditions of no overall deficit in food availability, to those who clearly need it.

No doubt such cases have occurred, although their frequency in the terms so described may be disputed. However, this is not a criticism of the Invisible Hand as such, but of the property rights structure within which it operates; as I suggested earlier, a variety of moral considerations may be brought to bear on the question of property entitlements. The apparent failure of a spontaneous market to distribute food to the needy may be the result of some previous violent acquisition of land and resources by a minority.

---

[21] Sen, "Profit Motive," p. 12. Bastiat's famous description of the state as "that great fiction by which everybody tries to live at the expense of everybody else" is apposite here.

[22] Ibid., pp. 10–11.

But even if it were the case that perfectly just acquisitions of resource holdings led to mass starvation in circumstances of aggregate adequate food supply, this does not refute the case for the Invisible Hand or establish the case for a *dirigiste* production and distribution system. Just because there have been occasional examples of such suffering under capitalist systems, it does not follow necessarily that collectivism itself would have prevented them. Surely, the more frequent cases of mass starvation under communism, as in the Soviet Union, are sufficient to tip the balance in favor of laissez-faire even in the least propitious cases.

This points to a more general problem with the approach of the critics of the Invisible Hand. Their obsessive concern with the alleged failures of the market process leads them to overlook the deficiencies of politics. All too often, the choice of methods is presented as if it were between the commands of a benevolent and omniscient legislator on the one hand and a messy and imperfect market on the other. However, the real comparison ought to be between political and economic means to generally agreed ends.[23]

Such a comparison is particularly instructive when we look at the results of competition for votes in a democracy. Under conditions of unlimited government, as in the United Kingdom, we find that the state, far from being the infallible purveyor of the public good, becomes the plaything of interest groups who use legal and political privileges to secure their members incomes they could not earn in the market. The tragedy here is that such action is ultimately destructive of the interests of all in the political community. Whereas the corrective mechanisms of spontaneous markets act to harmonize the self-interested actions of individuals with the common interest, this cannot be said of the political system. Thus, when Sen talks of the prevalence of prisoners' dilemmas in economic society,[24] that is, where self-interested action is ultimately self-defeating, he neglects to observe that their most obvious manifestations lie in the public world of political actors rather than in the private world of economic maximizers. Indeed, the increase in government activity during the last four decades has come about not through the need for collective action to produce public goods, as Sen claims, but through the democratic process itself.

[23] Samuel Brittan's work, especially *The Economic Consequences of Democracy* (London: Temple Smith, 1977), is most important in this respect.

[24] Sen, "Profit Motive," p. 18.

The Self-Interest Postulate

This brings us face-to-face with the problem of self-interest and the validity of the now well-known thesis of the late Fred Hirsch that the classic motivational postulate of the Invisible Hand theorem is an inadequate mechanism for satisfying everyone's interests. Hirsch made much of the phenomenon of "positional goods," by which he meant those goods, such as rare works of art or uncrowded beaches, whose supply cannot increase with increase in demand. He implies, in other words, that a market economy produces a mass of unsatisfiable desires. Perhaps the novelty of the concept has been exaggerated. Positional goods have always been with us and, presumably, always will; their existence does not point to the failure of the Invisible Hand but to unalterable facts of the world. It seems to me that Hahn[25] has it about right when he suggests that the market will always provide an adequate supply (and immense variety) of augmentable goods, which makes the positional goods problem less pressing than is often supposed. Certainly, Hirsch's implicit claim that the only solution is to alter somehow individual wants has an unduly depressing (and faintly sinister) ring to it.

There is, however, in Hirsch's work a more direct argument that the pursuit of self-interest is counterproductive, that rational individuals cannot achieve their undeniably selfish ends by the pursuit of self-interest alone. It is claimed that certain kinds of nonselfish behavioral motivations, such as "team spirit" or "loyalty," both of which may require individuals to sacrifice immediately beneficial opportunities, are required in order to generate the efficient outcomes that are supposed to be a product of unalloyed egoism. Sen[26] cites the Japanese commercial success as partly due to a peculiar ethos that has little to do with the individualistic spirit of Adam Smith.

The attack on self-interest, however, neither blemishes the Invisible Hand theory nor affects drastically the self-interest assumption frequently used by expositors of that theory. A universal self-interest is not a logically necessary feature of that theory, which merely claims that a social order will emerge from the spontaneous actions of individuals without the necessity for an all-powerful central institution. The importance of the self-interest proposition is to show how the public interest is an accidental and unintended outcome of private actions so that the appeal to a deliberate altruism is either redundant or positively harmful. As Adam Smith remarked, "I have never known much good done by those who affected to trade for the public good."

---

[25] Hahn, "Reflections on the Invisible Hand," p. 9.

[26] Sen, "Profit Motive," pp. 14–15.

General reliance on the postulate of self-interest to explain economic behavior, however, does not exclude the possibility that more efficient orders may be produced by arrangements that rely on other motivations. A natural process will select out in an evolutionary manner the successful industrial forms, and these may well include some that are characterized by nonegoistic motivations. The organization of the firm is a response to the uncertainty of economic life, and we cannot know in advance of experience which is the most efficient mode of its operation. Certainly, we cannot use the examples of enterprises that depend less on self-interest and are still successful as a license to inculcate *deliberately* a nonegoistic spirit. Where this has been tried the results have been disastrous. The point overlooked when examples of successful "team spirit" enterprises are cited is that the institutional arrangements and psychological attitudes that are said to be the causal factors in these success stories have developed spontaneously. Where they have not so developed the traditional concept of man as a maximizer is still the only serviceable economic notion that we can have.

In fact, few of the Invisible Hand theorists ever supposed that a society could function entirely by selfish impulses. The prevalent view has been that although the coherence of the market order could be explained by this postulate, the social and legal context in which trading takes place depends on a measure of self-restraint and moral probity that appears to be excluded by the psychological assumptions of traditional theory. But, of course, such behavioral traits are not excluded. All that is excluded is the construction of rationalistic moral codes that presuppose that man can be other than he is or declare that a new altruistic man, unfettered by the ethics of greed, will emerge if only capitalist institutions can be overturned in a revolutionary manner.

There is no externality problem here—the market has not failed to produce the public good of morality and self-restraint. Laissez-faire economic arrangements are quite consistent with the traditional Western ethical values of honesty, fair-dealing, justice, and individual autonomy. The nonobligatory virtues of benevolence and charity flourish under capitalist systems. There is evidence of much greater selfishness and brutality in socialist regimes precisely because the market—the one area where egoism can beneficially materialize itself—has been suppressed.

Flaws in the Negative Externality Argument

In comparison with these rather grand themes, the other areas of alleged market failure, where the Invisible Hand has failed to steer society in a smooth manner, seem almost routine and mundane. These cover the whole range of negative externalities, where private transactors impose costs on the community as a whole. This is now such a well-worn theme that a detailed summary would not be relevant here. It is sufficient to point out that the original argument (from Pigou) that an observed divergence between private and social costs automatically negated the case for the Invisible Hand and justified government intervention is no longer as plausible as it once was. This is so on at least two grounds. First, as mentioned earlier, it must not be assumed that government is an omniscient and benevolent institution that automatically maximizes the public interest. In fact, political failure is likely to be more common than market failure because political actors operate (normally) under fewer constraints than economic transactors. The public choice theorists have strengthened the case for the Invisible Hand with their systematic demonstrations of the failure of government, although this was probably not their deliberate intention.

Second, the rapidly expanding property rights literature has shown that many of the alleged externality problems could be internalized if the ownership of resources could be specified. In an appropriate legal framework, there would be many possibilities for trade between transactors in the face of apparent externalities. In many cases, there are already appropriate institutional arrangements so that the need for government to conjoin private and social costs by coercive taxation or prohibition is otiose. The common law itself contains remedies for people adversely affected by the economic actions of others.[27]

Conclusion

Even these brief considerations suggest that economists should be more concerned with showing how institutional re-arrangements might assist in the working of the Invisible Hand rather than demonstrating gleefully how far some existing market allocation departs from an alleged optimum. If there is no Invisible Hand process then there are no economic regularities, and if there are no such regularities there is no economic science. In the absence of the Invisible Hand, the economic world becomes the plaything of, first (and very briefly), the delicate and elegant but headily unrealistic abstract theorist, and then the defenseless target of the clenched fist of the political commissar.

---

[27] The classic work here is Ronald H. Coase, "The Problem of Social Cost," *Journal of Law and Economics* 3 (October 1960): 1–44.

END THE POSTAL MONOPOLY

James C. Miller III

Earlier this year, the cost of mailing a first-class letter went from 20 cents to 22 cents. It was the third increase in four years, the seventh since the Post Office became the independent U.S. Postal Service in 1970. As usual, politicians decried the inefficiency of postal delivery and promised to "do something." But few seemed willing to consider the one action likely to have a real effect on the efficiency of the U.S. postal system: let others compete in the delivery of first-class mail.

The Evolution of the Postal Monopoly

While the Constitution provides Congress with the power "to establish post offices and post roads," it does not require that the carriage of mail be a monopoly, much less a government monopoly. Throughout the first half of the 19th century, however, Congress enacted ever more restrictive measures limiting the private carriage of mail. For example, an 1825 statute that reserved to the federal government the exclusive right to carry letters for hire in vehicles over post roads was amended two years later to prevent private carriage on horseback or foot. This presumably made life difficult for the Pony Express, which although often used as a symbol of U.S. mail service, was a creature of private enterprise.

Finally, Congress passed the Postal Act of 1845—the so-called private express statutes—which directly prohibited private companies from carrying letters for hire.[1] These statutes (with a few modifications) remain the law today. They grant the government a monop-

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Chairman of the Federal Trade Commission. The views expressed here are his own and do not necessarily reflect those of the other commissioners. Portions of this article are drawn from his testimony before the Subcommittee on Economic Goals and Intergovernmental Policy of the Joint Economic Committee, 21 June 1982.

[1] 5 Stat. 732 (3 March 1845).

oly in the delivery of letters (first-class mail). The monopoly does not extend to newspapers and magazines (second-class mail), direct-mail advertising (third-class mail), or parcels (fourth-class mail).

Through its ability to define a "letter," the Postal Service is in the enviable position of being able to determine the extent of its own monopoly. While the service has "suspended" its monopoly for certain letters (such as time-sensitive materials), it has also expanded its monopoly by defining letters to include bills, receipts, IBM cards, magnetic cards, magnetic tapes, and other business documents.[2] Typically, as new services such as express mail have developed, the Postal Service has first asserted that these services fall within its monopoly and then announced a suspension of the monopoly with respect to *some* aspects of the new service.[3]

Is the Postal Monopoly Economically Justified?

The standard economic rationale for production by a regulated franchise monopoly, whether privately or publicly owned, is that, under certain technical conditions, production may be accomplished at least cost by a single firm. Government may wish to impose maximum rate-of-return regulation on such "natural monopolies" in an attempt to reduce the allocative efficiency costs associated with unregulated monopolies. But even in the case of a proven natural monopoly, economists are ambivalent about the desirability of regulation because of its direct and indirect costs.[4]

There is, however, no convincing evidence that mail service is a natural monopoly. Indeed, most of the relevant studies have failed to find significant economies of scale.[5]

---

[2] The USPS currently defines a "letter" as any "message directed to a specific person or address and recorded in or on a tangible object." 39 C.F.R. §310.1(a).

[3] The USPS has asserted that it has authority to exempt materials from the postal monopoly pursuant to 39 U.S.C. §601(b).

[4] See, for example, Paul L. Joskow and Roger C. Noll, "Regulation in Theory and Practice: An Overview," in Gary Fromm, editor, *Studies in Public Regulation* (Cambridge: MIT Press, 1981), p. 10.

[5] For instance, an empirical study by the Post Office itself did not uncover economies of scale. See Bureau of Finance and Administration, U.S. Post Office Department, *Summary Report of Cost System Task Force on Incremental Costs* (Washington, D.C., May 1970). See also the review of studies in Leonard Merewitz, "Economies of Scale in Postal Service" (U.S. Postal Rate Commission (PRC); 30 August 1973) and the review contained in Initial Decision, PRC Docket No. R74-1, pp. 78–90. No general economies of scale were found in Rodney E. Stevenson, "Postal Pricing Problems and Production Functions" (Ph.D. dissertation, Michigan State University, 1973), and the Department of Justice found "virtually no credible evidence that this [i.e., economies of scale] is actually the case." See U.S. Department of Justice, *Changing the Private Express Laws* (1977).

Most importantly, if mail service were a natural monopoly, competitors would find it very difficult to provide the same level of service at the same or lower cost.[6] Consequently, the Postal Service's position would be secure without legal protection.

The Economic Costs of the Postal Monopoly

The Postal Service provides a classic illustration of the problems inherent in trying to determine efficient prices and hold down costs for a regulated monopoly.

The existence of large common costs in enterprises like the Postal Service makes it impossible to allocate total costs to individual services in a nonarbitrary manner.[7] This makes ratemaking very difficult. While it is conceptually possible to determine efficient prices by looking at both cost and demand conditions, the necessary information is typically not available in sufficient detail. As a result, there is no reason to believe that prices will be efficient—even before allowing for the influence of a variety of political pressures that make themselves felt during the ratemaking process.

Moreover, with restricted entry, an inefficient rate structure is easy to maintain, because new firms cannot enter in response to a price that is set too high. Thus, cross-subsidization, which is virtually impossible under competitive conditions, occurs with some frequency in regulated industries. It is, in fact, a major source of the economic inefficiency associated with the Postal Service. In addition, postal costs are generally higher because the Postal Service faces only limited incentives to produce in the least-cost manner.

---

[6] Under special circumstances, it is theoretically possible for an entrant to provide one of a natural monopolist's services at a price lower than the most efficient price, in which case the natural monoply is said to be "unsustainable." See John C. Panzar and Robert D. Willig, "Free Entry and the Sustainability of Natural Monopoly," *Bell Journal of Economics* (1977), pp. 1–22. But we have no evidence that the postal monopoly is unsustainable. In fact, "there is no evidence that the cost conditions for that are present in postal services." See Bruce M. Owen and Robert D. Willig, "Economics and Postal Pricing Policy," in Joel L. Fleishman, ed., *The Future of the Postal Service* (New York: Praeger Publishers, 1983), p. 231. More generally, the sustainability literature is "too specialized and insufficiently operational" to support entry restrictions as a policy prescription. See Joskow and Noll, p. 17.

[7] See Melvyn A. Fuss, "Cost Allocation: How Can the Costs of Postal Services Be Determined," in Roger Sherman, ed., *Perspectives on Postal Service Issues* (Washington: American Enterprise Institute, 1980), pp. 30–46; Owen and Willig, pp. 227–45; and Rodney E. Stevenson, "The Pricing of Postal Services," in Harry M. Trebing, ed., *New Dimensions in Public Utility Pricing* (East Lansing: Michigan State University, 1976), pp. 427–52.

Inefficient Rate Structure

Because of the problems involved in allocating costs, precise quantitative estimates of cross-subsidies within the postal system are difficult to obtain. It is generally conceded, however, that high-density urban service subsidizes low-density rural service. It is also alleged that first-class mail in general subsidizes some or all of the other classes.

The subsidy from urban to rural users arises primarily because the Postal Service charges the same rate for all first-class mail, regardless of the distance the mail travels and the costs of delivering it. Some first-class mail is therefore overpriced and some is underpriced. Consequently, as the Postal Service itself has stated, "if the Private Express Statutes were repealed, private enterprise . . . would be free to move into the most economically attractive markets. . . ."[8]

One of those areas would undoubtedly be transaction mail—bills, bank statements, and so forth. This mail, which tends to be generated by a few, high-volume local business mailers (for example, banks, utilities, and department stores) and tends to be destined for high-density areas, is relatively inexpensive to deliver.[9] If entry were permitted, the price for delivering transaction mail (which constitutes a significant portion of first-class mail) would fall. This is what the competitive process is all about, notwithstanding the fact that the Postal Service refers to it as "cream-skimming."[10] Such cream-skimming opportunities only occur when prices exceed costs. The existence of such opportunities is therefore evidence that current pricing misallocates resources.[11]

Whether there is an additional subsidy going from first-class mail to other classes is an issue of some contention. However, the Postal Rate Commission's chief administrative law judge concluded that such a subsidy did exist, and that "the Postal Service has become a tax-collecting agency collecting money from first-class mailers to distribute to other favored classes."[12] If this is the case, entering firms would be able to lower the average price charged for first-class mail.

---

[8] U.S. Postal Service, Statutes Restricting Private Carriage of Mail and their Administration (House Committee on Post Office and Civil Service, 93d Cong., 1st Sess., Committee Print 1973), p. 6.

[9] The Postal Service offers discounts for presorted first-class mail of 500 or more pieces. It also offers discounts for first-class mail addressed with machine-readable "Zip + four" address labels. However, these discounts do not take into account the number of separate places to which the mail is addressed, nor the population density of those locations.

[10] Ibid., p. 5.

[11] See, for example, Alfred E. Kahn, *The Economics of Regulation: Principles and Institutions, Volume II: Institutional Issues* (New York: John Wiley and Sons, 1971), pp. 220–46.

Higher Overall Costs

The postal system is a particularly good source (in the event that one is needed) for evidence that private enterprise performs better than government enterprise and that competitive markets perform better than monopolies.[13]

The costs of the Postal Service are significantly higher than they should be because the incentive to hold down costs—most notably, labor costs—is limited. Postal workers are paid far more than is necessary to retain their services.[14] Because entry into postal markets is restricted, the Postal Service is able to pass those higher costs along to its customers.

Higher costs also result from failure to innovate. For example, as of 1974, the Postal Service still shipped all of its parcels in bags, which resulted in high breakage rates and handling costs. When the United Parcel Service (UPS)—a private competitor—introduced mechanization and containerization, breakage was reduced and handling costs lowered. As a result, the Postal Service has steadily lost business to UPS.[15]

---

[12] See James C. Miller III and Roger Sherman, "Has the 1970 Act Been Fair to Mailers?" in Sherman, *Perspectives*, pp. 64–65.

[13] There are, of course, many other sources. See Roger Ahlbrandt, "Efficiency in the Provision of Fire Services," *Public Choice* 16 (1973): 1–15 (finding private firm provides goods at lower cost than government agency); W. Mark Crain and Asghar Zardkoohi, "A Test of the Property-Rights Theory of the Firm: Water Utilities in the United States," *Journal of Law and Economics* 21 (1978): 395–408 (finding higher operating costs in publicly owned water utilities); David G. Davies, "The Efficiency of Public v. Private Firms: The Case of Australia's Two Airlines," *Journal of Law and Economics* 14 (1971): 149–65 (finding private company more efficient than public firm); Cotton M. Lindsay, "A Theory of Government Enterprise," *Journal of Political Economy* 84 (1976): 1061–77 (determining government managers maximize self-interest by producing output Congress is likely to value more highly than consumers do); and Sam Peltzman, "Pricing in Public and Private Enterprises: Electric Utilities in the United States," *Journal of Law and Economics* 14 (1977): 109–47 (finding that government-owned firms adopt pricing policies that enhance their political support, not consumer welfare).

A number of Federal Communications Commission (FCC) decisions that altered the monoply conditions formerly prevailing in the common carrier communications field provide empirical evidence that competitive markets perform better than monopolized ones. The Chairman of AT&T, for example, testified that as a direct result of the FCC's rulings, the Bell System companies were forced to become more innovative and to deploy the results of that innovation more rapidly. In addition, the prices Bell charged were generally reduced and new policies reflecting more cost-related pricing were adopted. See generally letter from FCC Chairman R. E. Wiley to All Members of Congress on H.R. 12323, 25 May 1976; and FCC, Second Report in Docket 20003, FCC 80-5, released 29 January 1980.

[14] See Douglas K. Adie, "How Have Postal Workers Faired Since the 1970 Act?" in Sherman, *Perspectives*, pp. 74–93.

The success of private express mail service is also indicative of the benefits of competition. Despite the threat of suit by the Postal Service, numerous private carriers developed in the 1960s and 1970s, specializing in the rapid and reliable transportation of commercial documents. In 1973 the Postal Service responded and established its own express delivery service called "Express Mail." Despite attempts to expand the postal monopoly to include express mail services, private services now compete openly and freely. Indeed, private companies such as Federal Express and Purolator deliver overnight to more cities than does the Postal Service.

Fears of Higher Prices and Reduced Service

If repeal of the private express statutes will reduce rates for some postal services, will it also raise rates for others? This is a major fear, particularly for residents of rural areas. There are, however, two reasons for believing that such fears are exaggerated.

First, if competition reduces the overall level costs, this will benefit everyone, including consumers of high-cost services. Second, our experience with the deregulation of trucking and airlines suggests that fears of significant reductions in rural service due to postal deregulation are probably unfounded. Some small cities are no longer served by large jets, but commuter airlines have used smaller planes to serve small communities at far lower costs. Thousands of new companies have gone into trucking, and as a result trucking service to rural areas has improved.

If we can obtain airline and trucking service throughout the country without a government monopoly, competition can also work for letter delivery. The configuration of service may change somewhat, but surely everyone would have access to affordable mail delivery.

If a subsidy is deemed necessary, it might be patterned after the small-community air service program implemented along with airline deregulation. This program now subsidizes air service to 145 communities at an annual cost of $51 million. This cost has declined significantly since the program began in 1978, and the program is scheduled to be phased out. Yet, on the whole air service to small communities has improved since deregulation.

---

[15] The USPS tried to learn from UPS's success by building a one billion dollar bulk mail system to improve its handling of parcel post mail. But that effort has been such a failure that the General Accounting Office has suggested USPS write the investment off as a loss. See, for example, Joel L. Fleishman, "Postal Policy and Public Accountability: Is the 1970 Bargain Coming Unglued?" Manuscript, Harvard University, Program on Information Resources Policy, 1981, p. 94.

The cost of the small-community air service program has been running at about 0.2 percent of total domestic airline revenues. The analogous figure for the Postal Service would be $26 million—a paltry sum for ensuring that there will be universal mail delivery, and for enabling the benefits that competition would bring to this industry.

Conclusion

The burden of showing that the postal monopoly is necessary or desirable has not been met. All the available evidence suggests that competition in the market for first-class letter delivery would create substantial benefits.

If the private-express statutes were repealed, existing mail firms would expand and new ones would enter the market. Probably just as important, the spur of competition would mean improvements in the Postal Service itself. Private enterprise will get the mail delivered—just as it did in the Old West.

IS SOCIAL SECURITY POLITICALLY UNTOUCHABLE?

Loren E. Lomasky

I. Introduction

In recent decades Social Security has been a fount of periodic fiscal crises, each more severe than the previous. Within the economics profession there has been no shortage of proposals for the reform or radical restructuring of the system, and in this paper no further recipes shall be added to a cookbook already bulging with ingenious concoctions. Instead, I wish to focus attention on what might be the most remarkable feature of all these proposals: their political irrelevance. Despite grave intonations in professional journals and the media that Social Security is sorely troubled, there exists no constituency whatsoever for anything more than tinkering with the retirement provisions of the current system. Indicative of the high degree of resistance to fundamental restructuring of those provisions is the statement of the National Commission on Social Security Reform:

> The members of the National Commission believe that the Congress, in its deliberations on financing proposals, should not alter the fundamental structure of the Social Security program or undermine its fundamental principles. The National Commission considered, but rejected, proposals to make the Social Security program a voluntary one, or to transfer it into a program under which benefits are a product exclusively of the contributions paid, or to convert it into a fully-funded program, or to change it to a program under which benefits are conditioned on the showing of financial need.[1]

This strong endorsement of the status quo was offered, without dissent, by a broad-based committee called to deal with the most

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Associate Professor of Philosophy at the University of Minnesota at Duluth. He wishes to thank Geoffrey Brennan, James Buchanan, James Dorn, and Richard Wagner for helpful suggestions that aided the writing of this paper.

[1] *Report of the National Commission on Social Security Reform* (1983, chap. 2, p. 2).

heralded financial shortfall yet to be encountered by the Social Security system. Politicians have uniformly recognized that to impugn the system's desirability, or even to raise the question of whether an alternative means of providing retirement income might be preferable, is to commit political suicide. Approval of Social Security is, in politically relevant terms, unanimous.

Why this should be so is mysterious. The retirement provisions of Social Security constitute an intergenerational transfer mechanism that redistributes income from the young to the old.[2] It is not at all surprising that the latter group supports the provisions of the current system, but an explanation is called for as to why others are either supportive of Social Security or, at most, indifferent. The financial burden on those who are net losers is not trivial: More than one-third of all households in the United States pay more in Social Security taxes than they do in federal income tax. This then is not a case in which small impositions are spread out across a large number of persons for each of whom the burden is too slight to be noticeable. Nor is it the case that support for Social Security by the young is explicable by any conviction on their part that the present value of their lifetime payments into the system is exceeded by the present value of the benefits they expect to receive. Private pensions offer the young a higher rate of return than does Social Security, an imbalance accentuated by the 1983 reforms. Even more significant are results from national polls indicating that as many as 84 percent of those between the ages of 18 and 44 do not believe that promised benefits will continue to be available when they retire.[3] Moreover, it is likely that a large number of those who do expect to receive some retirement benefits from Social Security expect that the amount they eventually receive will be less than that paid in. These considerations appear to be a prescription for intergenerational warfare with the young assaulting the foundations of Social Security and the old resolutely defending them. But no such battle has commenced, nor are there signs of one in the offing.

---

[2] That is not to deny the existence of other significant transfer effects, for example, from upper- to lower-income groups or from females to males. Rather, it is to emphasize the preponderance of the intergenerational transfer effect.

[3] For example, a CBS News/New York Times poll of June 28–July 1, 1982 asked: "Do you think the Social Security system will have the money available to provide the benefits you expect for your retirement?" Of those respondents who expressed an opinion, 80 percent of those aged 18 to 29 and 76 percent of those aged 30 to 34 responded in the negative. Even for those aged 45 to 54, a small majority answered no. Surveys conducted on different dates and with variously worded questions have elicited similar responses.

Were there an actual or latent constituency for the abolition or radical revision of Social Security, one would expect that political entrepreneurs would seize the opportunity to mobilize it. Conversely, in the absence of any attempt to build a political coalition based on a platform of opposition to Social Security, it is necessary to conclude that no such constituency exists. What requires explanation is why it does not, why Social Security is politically untouchable.

II. The Economic Theory of Politics: A Reexamination

The unchallenged status of Social Security is an embarrassment for the economic theory of politics.[4] If individuals vote their self-interest and if politicians position themselves to be the recipients of self-interested voting, then the political waters ought to be mightily roiled by Social Security. Yet they remain pacific. Various reasons for this calm have been suggested: fiscal illusion, public goods aspects of Social Security, informational communication failures, and naivete.[5] It has, in addition, been regularly predicted that although no significant opposition to Social Security has *yet* surfaced, such resistance is brewing and will soon erupt.[6] These attempts to protect a favored theory from disconfirmation seem ad hoc. Empirically better grounded is Rudolph Penner's assertion: "Anyone who believes that radical reforms of the Social Security system are possible is already living in an imaginary world" (1983, p. 465).

If the theory of self-interested voting is to be regarded as a genuinely empirical hypothesis subject to falsification, then Social Security must be acknowledged to be a crucial test of its warrant. The pecuniary stakes are high, net gainers and net losers are distinguishable with a high degree of precision,[7] and vast amounts of information—and misinformation—have bombarded the public. Moreover, the system has operated in its current form for several decades,[8] long enough, one would think, for nascent opposition to have coalesced into a politically potent coalition. One is hard pressed to identify any other national issue for which the conditions favoring informed, rationally self-interested voting are so well satisfied.

---

[4] It is admittedly an oversimplification to speak of "the economic theory of politics" as if it constitutes one precisely specified account to which nearly all analysts subscribe. My understanding of the literature is that there is a dominant tendency shared by many theorists who nonetheless differ significantly in the manner of application. That tendency is to assume (i) that voting is an instrumental exercise in which voters choose a candidate or policy based on expected returns from a vote; (ii) that the expected return can be entirely or almost entirely explained in terms of narrow self-interest understood as economic well-being; (iii) that politicians assume policy positions of a form that enhances their chance of capturing such self-interested votes. Among the more important places in which such a view is developed are Downs (1957), Buchanan and Tullock (1962), Riker and Ordeshook (1973), Stigler (1973), and Tufte (1978). The model of self-interested rationality that the economic theory of politics presupposes has occasionally been subjected to criticism, but because no comparably powerful conception has been offered by way of replacement, such criticism has had little effect. See Edelman (1964) and Hartwig (1981).

[5] See, for example, Browning (1975), Mitchell (1977), Patton (1977), and Browning (1979).

[6] See Weaver (1982).

Three avenues are available for explaining the theory's apparent breakdown. First, one can point to the continued reliance on ad hoc devices, a recourse not as methodologically suspect as may at first seem to be the case. Philosophers of science have demonstrated that an otherwise well confirmed theory can withstand the existence of perplexing anomalies if there is no equally well confirmed alternative theory that can account for the troublesome instances.[9] Resort to ad hoc devices to explain Social Security's political standing has been the approach taken by most economists.

Second, one may reject a *private-interest* account of voting in favor of a *public-interest* theory. It is clear that a *public-interest* hypothesis fits the case of Social Security far better than do standard models of self-interested voter behavior. However, to adopt such an approach as this would be to abandon the analytical gains that have been realized by the economic theory of politics. It is retrogressive to suggest that the fruits of that theory are illusory and that we must, after all, return to a model that has been regarded as inadequate for the past quarter-century.[10] A less heroic variant would be to maintain that self-interest sometimes explains the behavior of an electorate, while in other circumstances public-interest concerns are dominant. An explanation would still be needed, however, concerning why Social Security falls so decisively into the latter class.

---

[7] The dividing point will shift with changed legislative provisions and as economic and demographic factors vary. It is not being asserted that the average voter can readily calculate the economic impact Social Security has on him. However, one would surmise that if voters can be counted on to vote their self-interest, political entrepreneurs ought to have an incentive to see that the relevant calculations are performed and to bring them to the attention of net losers. Some results are presented by Ferrara (1980), especially pp. 112–74; Munro and Parsons (1977); and O'Neill (1979). Calculations exhibiting the increased losses to younger persons occasioned by the 1983 Social Security legislation can be found in Goodfellow and Pellechio (1983).

[8] The Social Security Act of 1935 created a funded pension system. Amendments in 1939 began the change to the current pay-as-you-go scheme.

[9] See Kuhn (1970 and 1977). Some significant revisions of the Kuhnian thesis are offered by Laudan (1977). A lucid work that is more broadly epistemological is that of Quine and Ullian (1978).

[10] Use of Ockham's Razor suggests that an account that models the motivation of agents' political behavior on presuppositions fundamentally different from agents' market behavior is prima facie suspect. Only if no satisfactory unitary theory can be devised to explain both will it be acceptable to introduce such a dichotomization of behavioral motivation. An attempted resurrection of the public interest theory of voting behavior is offered by Kiewet and Kinder (1979).

Third, one can retain the hypothesis that individuals do vote their self-interest, but modify the standard account of the nature of self-interested voting. This is the approach taken in the present paper. It will be shown that too narrow a view of the parameters of self-interest has infected the economic analysis of political behavior. By and large, theorists have been so enamored of the possibility of applying considerations bearing on market choice to instances of public choice that they have failed to see important dissimilarities. In particular, voting behavior is typically *inconsequential* in a way that market behavior is not. Because of the relative inconsequentiality of voting, individuals in the polling booth regularly have an incentive to lend decisive weight to considerations that are usually submerged in market choice. In other words, an individual will vote for policies that he would reject were his choice decisive. Moreover, he will do so in the strict pursuit of his rational self-interest.[11]

Politicians are also to be conceived of as rational maximizers; their goal, to oversimplify in a harmless way, is to maximize their vote totals. It may be surmised that successful politicians have recognized a fact that has escaped theorists: self-interested voting can and does diverge from voting for the policy that would, if adopted, most advance the voter's welfare. Accordingly, when such divergence is exhibited, politicians will regularly support policies designed to capture the *votes* of the majority, rather than support policies that would enhance the *welfare* of the majority.

The analytical structure supporting this revision of the self-interested voter model will be set out in sections III and IV, and its application to the politics of Social Security will be developed in section V. In section VI, I conclude with observations concerning those conditions under which significant opposition to Social Security *could* be expected to arise. It may, after all, prove to be the case that the untouchability of Social Security is only a near- and intermediate-term phenomenon.

III. Market Choice versus Political Choice

A significant characteristic of market choice is that an individual who acts in the market is, typically, *decisive*. If he opts to purchase

---

[11] This account is further developed in Brennan and Buchanan (1982) and Brennan and Lomasky (1982; 1983).

A rather than B, he receives A and incurs the opportunity cost of the forgone B. It follows that market choice is *consequential* in the sense that an individual acts in the knowledge that he will enjoy the benefits and bear the costs of the choices he makes (positive and negative externalities excepted). In such a circumstance, it is entirely reasonable to suppose that an individual will select A over B if and only if it is in his perceived self-interest to do so.

In contrast to market choice, it should be immediately apparent that an individual's vote, when faced with the choice of voting for policy A or policy B, will be decisive only when the remainder of the electorate is evenly split between A and B. If the number of other voters is 2n, then the probability of a tie occurring is the number of ways in which 2n voters can be arrayed such that A receives exactly n votes, divided by the total number of ways in which 2n voters can be arrayed, that is:

$$p = \frac{{}^{2n}C_n}{2^{2n}} \tag{1}$$

Using Stirling's approximation when n is large and when the prior probability that a random voter will vote for A is 0.5, the corresponding approximation to (1) is:

$$p = \frac{1}{\sqrt{\pi n}} \tag{2}$$

This represents the highest probability that an individual's vote will be decisive in an electorate of 2n + 1 independent voters. If a random voter's likelihood of voting for A is more or less than 0.5, then the probability of an exact tie among 2n voters is less than (2).

Even if we ignore the problem of specifying more finely the prospects of an individual voter casting the swing vote, it is clear that when the electorate is even moderately large, the probability of one's vote being decisive under conditions of majority rule is small.[12] A rational voter cannot be supposed to be oblivious to this fact.

If we let $V_i(A)$ represent the value the ith voter attaches to the enactment of A, and $V_i(B)$ represent the value assigned to the enactment of B, the expected return on a vote for A is:

$$R_i = p[V_i(A) - V_i(B)] \tag{3}$$

Even if alternative A is strongly preferred to B, the expected return on a vote for A may be very small if the electorate is large. At times, therefore, it will not be costly to *vote* for B rather than A, even if the actual *imposition* of B would impose substantial costs. This is the meaning of the assertion that electoral choice is relatively inconsequential compared with market choice.

---

[12] The relevant calculations are set out more elaborately in Beck (1975).

If an individual's vote is very unlikely to be decisive, it seems a misuse of language to refer to electoral determinations as instances of "public choice." No voter *chooses* an outcome in the sense of *bringing about* that outcome in the way that an individual exhibits choice in the market. To be more precise, a voter does not choose policies or candidates; he chooses only which lever to pull. Thus, if there exists some *direct payoff* from a vote for B rather than A, a self-interested voter may find it advantageous to vote for B even though R_i is positive. In such a case, the ith voter's decision to vote for B is rational, even though the adoption of B would substantially impair his welfare.

These considerations are expressed schematically in Figure 1. If the direct payoff from a vote for B is $P_1(B)$, then a voter will maximize the return on his vote by voting for A when the size of the electorate is less than Y, but by voting for B when the electorate is greater than Y. If the direct payoff increases to $P_2(B)$, then the electorate need only be greater than X to render a vote for B in the ith voter's self-interest.

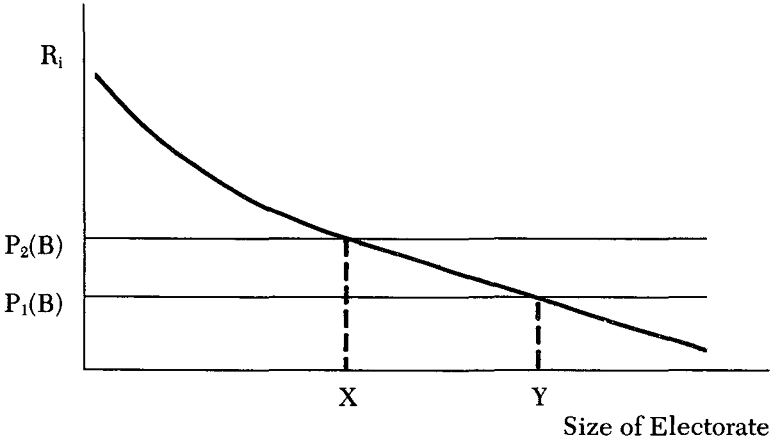


FIGURE 1

EXPECTED RETURN ON A VOTE FOR A

No content has heretofore been given to the concept of a direct payoff from a vote for B. One obvious candidate is a bribe. If the electorate is greater than Y, the voter will be acting in his self-interest to accept a bribe of $P_1(B)$ to cast a ballot for B. His "services" will be more expensive if the electorate is smaller.[13] Elections in which bribes or side-payments are possible are those in which one can expect to find votes deviating from policy preferences.

Although the economic theory of bribery as applied to electorates of various sizes has some inherent interest, its general applicability to majoritarian voting is slight.[14] Bribes were introduced simply to provide one clear instance of direct payoff from a vote. It will now be argued that other such direct payoffs exist—payoffs less tangible but more commonly encountered than bribes—and that they provide a means for explaining the otherwise puzzling political status of Social Security.

IV. The Importance of Expressive Returns

Theorists have long recognized that it is problematic why wealth-maximizing individuals bother to vote in large-number elections where the value of R_i must be very small.[15] The costs in expended time and energy exceed whatever return is expected to accrue to a vote. The acknowledgment of this difficulty, often reluctantly made, has not led to a significant revision of the economic theory of political behavior, however. It has sometimes been supposed that some public-interest motivation accounts for the fact that individuals *do* vote, but that once they are in the polling booth economic interest determines *how* they vote.

Such a schizophrenic account merits rejection. It is without warrant to maintain that the motives that draw an individual to the polls suddenly disappear once the curtain is closed and he confronts rows of levers. A prima facie acceptable model must present a uniform motivational account; if self-interest is taken to explain the direction of voting, then it should also have bearing on the occurrence of voting. Recognition that a voter receives a direct payoff from casting a ballot satisfies this uniformity requirement. Voting can thereby be understood as not merely instrumental, calculated to increase the probability of victory of the favored alternative, but also as a consumption activity affording a direct payoff to the voter. A direct payoff would explain why voters incur costs to transport themselves to the polls; it would also help explain why they vote as they do (just as a bribe can account for both the fact of voting and the nature of a vote cast).

---

[13] I abstract from the briber's problems of ensuring compliance on the part of the bribee.

[14] Or is this the pious hope of an author suffering from a surfeit of idealism?

[15] See Downs (1957). Ferejohn and Fiorina (1974) have addressed this apparent irrationality, proposing a "minimax regret principle" by way of explanation. That rather farfetched device is rendered superfluous by the model constructed in this paper and in the Brennan-Buchanan (1982) and Brennan-Lomasky (1982; 1983) papers.

The consumption goods that voting provides can be various, and they might be idiosyncratic. No attempt will be made here to catalog the range of possible direct payoffs. Rather, attention will be directed to one consumption feature of voting that is liable to be widespread, that can prominently influence results in large-number electorates, and that strongly suggests itself as a major factor underlying the political untouchability of Social Security. This feature will be referred to as the *expressive return* from a vote.

By "expressive return" is meant those benefits derived from expressing support for an outcome, independent of whatever effect the expression of support might have on the likelihood that the desired outcome will emerge. Two instances will help clarify this concept.

On any given autumnal Saturday, hundreds of stadiums across America are packed with fans lustily cheering their favorite football teams. In doing so they express a desire for a favored outcome—the victory of their favorite teams. It is ludicrous, however, to suggest that the fans' behavior can be explained as a deliberate attempt to increase the probability of victory by their favorite teams, even if it transpires that fan support *is* a causally relevant factor in a team's success. In other words, even if cheering in the aggregate affects a game's outcome, individual fans are unlikely to view their cheering as decisive. Rather, they will consider their cheering to be largely expressive, that is, they will view it as overwhelmingly a consumption good. Thus, an individual who incurs significant expense to attend a game to cheer for his team (or lesser expense to cheer in front of his television set) is not thereby acting irrationally or acting other than in his own self-interest.

Consider the similar case of one who sends a get-well card to a sick friend. Such a person is expressing support for a particular outcome. It may be that the receipt of such a card has some discernible therapeutic effect, but one would not ordinarily attempt to explain the card sender's behavior as an attempt to produce that effect. Rather, card sending is a predominantly expressive act, valued because of its expressive features.

The old adage "talk is cheap" can be understood as a recognition of the fact that, in many circumstances, it is costly to act to bring about an outcome, but it is relatively inexpensive to express support for that outcome. Economists have generally recognized this fact and thus discount opinion poll results as less-than-reliable indicators of how individuals act in market situations where costs are significant. What has been less often noted is that other avenues of expression, including voting within a large-number electorate, are also virtually costless. For precisely the same reason that leads theorists to conclude that market behavior will diverge from responses to opinion polls, it is predictable that voting behavior will diverge from private activity. The divergence appears because the cost structures are different.

During our schooling and upbringing certain moral principles were instilled in us, principles that continue to influence our behavior even though they do not dominate day-to-day activity. For example, one generally accepted moral principle is that the financially well-off ought to give some share of their assets to those less fortunate.[16] Such giving, however, represents a very small percentage of most people's expenditures. The explanation for this fact is not elusive: private giving is costly. The opportunity cost of a dollar's worth of charitable relief is a dollar of private consumption forgone. Persons who acknowledge charity as a moral duty but who indulge it infrequently should not be dismissed as simply hypocritical.[17] They might sincerely accept a principle of charity but find its application too costly given their other priorities. However, if the cost of extending a dollar of charity dropped to a few pennies, they would find it in their overall interest to give.

Although acting to produce an outcome is costly, voting in favor of an outcome is virtually a free good. In large-number electorates where R_i is correspondingly minute, a vote dictated by allegiance to some principle can prove to be a bargain. There are psychic costs incurred by consistently neglecting in one's activity principles that, at some level, one believes to be morally mandatory. Psychotherapy might be able to relieve a person of such compunctions, but that too is expensive. It is far more cost-effective to channel one's principled tendencies into inexpensive expressive activity. If R_i is small, then voting for "moral" policy B may be a highly efficient course of action. One "does the right thing" by voting for B in a situation where a single vote is inconsequential. It does not follow, of course, that the person who votes for B would have chosen B over A were he in a position for his vote to be decisive.[18]

---

[16] Note that one who assents to this proposition is thereby engaging in expressive behavior rather than acting to bring about the outcome that is expressively supported. No economist would be surprised if the correlation between voicing support for charitable giving and the actual allocation of cash to charities were small.

[17] The definition of hypocrisy being assumed is: P acts hypocritically in expressing support for outcome X if and only if P would be unwilling to pay any sum, no matter how small, to bring about X rather than not-X. Problems of application arise when X is some past event or fictional event such that it is logically or otherwise impossible that P could act to bring about either X or not-X. One would have to revise the counterfactual such that the antecedent clause refers to what P *would be willing* to pay *were P in a position* to alter outcomes. Although the details of a precise statement are somewhat messy, the general point should be clear.

The present discussion sketches out an extension of the self-interested voter model in which expressive returns are explicitly included in the calculus of the rational voter. Let $P(B)$ represent the direct payoff from a vote for B. If $P(B) - R_i$ is positive, the ith voter will vote for B rather than A even if R_i is also positive. Moreover, if the value of $P(B) - R_i$ is greater than the costs incurred by the act of voting, then he has an incentive to appear at the polls. The same factor that explains the act of voting also explains its direction.

A rational politician, then, will try to secure a majority coalition by supporting B if $P(B) - R_i$ is positive for a majority of voters; if it is not positive, he will support A. The politician will be indifferent as to whether R_i is itself positive, for whether a particular policy would in fact advance the welfare of a majority of purely self-interested voters is not a politically relevant variable. What concerns the politician is the *number* of votes he receives, not the particular nature of the interest *motivating* those votes. A vote cast primarily to reap expressive returns counts exactly the same as one motivated by financial concerns.

V. Implications of Expressive Returns for the Untouchability of Social Security

It should be evident that Social Security is precisely the sort of issue for which expressive returns can be expected to be an important determinant of voter behavior. As a federal issue, the relevant electorates are all large: the nation, entire states, and congressional districts. The probability of being the swing voter in any of these is very small; therefore, the value of R_i is likely to be swamped in individual calculations by any expressive returns that may be present.

That substantial expressive returns are in fact present is also apparent. Obligations toward the elderly are acknowledged by most persons. Typically, such obligations are backed not simply by abstract principles mandating concern for those who are old, but also by personal relationships in which emotional intensity is great. This is a country where Mother's Day and Father's Day are sacred festivals in the civil religion, where grown children are expected to make the welfare of parents and grandparents a primary concern, and where conspicuous failure to do so is cause for opprobrium. Nor are these concerns the outcome of essentially external controls on behavior; obligations toward the elderly, especially near kin, are internalized. Individuals who perceive themselves as neglectful are apt to incur unpleasant feelings of guilt and remorse. Because emotional distress is itself a cost, rational persons will attempt to minimize it.

---

[18] Similar considerations are adduced by Tullock (1971).

In societies where the extended-family household is the norm, it can be expected that the urge to express concern for parents' welfare will be directed largely to private activity. But in a mobile society where children typically live apart from aged parents with whom they are not in constant contact, opportunities to express regard for parental well-being are more limited. That desire does not disappear, however, with diminished personal contact; its satisfaction is sought instead in the public sphere.[19]

To send monthly remittances to an aged parent or to provide him with living quarters within one's domicile decreases personal consumption. But voting for policies that transfer resources to the elderly costs almost nothing and yet provides considerable expressive returns. "I'm *not* neglecting the interests of the old," one can say—sincerely. The rationality of voting for the continued existence or even extension of Social Security benefits, of responding to national surveys by declaring one's fealty to the system, and so on, is not called into question by the possibility that all votes for Social Security, when aggregated, might entail policies that no single voter would have chosen were his the decisive vote. What would be plainly irrational would be for a voter or survey respondent to act as if his were the decisive vote when it is not.

It should not be supposed that expressive returns are available only to the young. Older voters too can avail themselves of principled reasons to back Social Security: They have *earned* the benefits that they now (or will soon) receive through their previous payments into the system; everyone has a *right* to a decent retirement income; they do not want to be a *burden* on their own children. Note that each of these justifications for supporting Social Security explicitly invokes some moral principle—and note how natural the application of each is to the discussion of Social Security. Of course, those who have reached retirement age or will soon do so also have a narrow financial reason to support the continued existence of the system. This does not mean, however, that expressive factors do not enter into their support, but rather that their votes are over-determined. Nor is it being claimed that rhetoric is more reliable in predicting how individuals act than is self-interest. Rather, what is being maintained is that an examination of rhetoric can help the theorist to determine when expressive benefits are predominant and thus when self-interest will be fueled largely by the opportunity to reap expressive returns.

---

[19] James Buchanan suggested in conversation an extension of this point. The initial generations who received Social Security benefits had their income flows and concommitant ability to save for retirement disrupted by the Great Depression. It might then seem especially appropriate to expressively support a transfer to those who, through no fault of their own, possessed low net wealth to fund their retirement years. Simple passage of time will render this factor progressively less important.

I have no credentials to present as an expert on voter psychology and so cannot claim any personal authority in validating the preceding motivational speculations. What can legitimately be cited, though, as confirmatory evidence is the behavior of those whose business it is to appraise correctly the motivations to which voters respond: professional politicians. It can be observed, almost without exception, that politicians not only support the continued existence of Social Security but couch their support in terms calculated to capitalize on expressive motivations. Moralistic proclamations are the common currency of their public remarks on Social Security, with objective discussion of the system's fiscal implications running far behind. Theorists and policy analysts may decry this phenomenon, but its significance as an empirical datum should not be neglected.

It is reasonable to maintain that successful political entrepreneurs supply the brand of rhetoric that voters demand. If voters demand opportunities to gain expressive returns through their votes, then politicians will comply. It misses the point to accuse politicians of demagoguery when they profess undying support for a system whose unfunded liabilities they simultaneously allow to grow to a level of some $7 trillion. They are merely acting in their own self-interest when they govern their activity to accord with the constraints of majority-rule institutions that strongly encourage expressive voting. (Or perhaps we ought to define "demagoguery" as the attempt to capitalize on expressive factors when doing so diminishes overall welfare. If so, then it must be acknowledged that democratic institutions offer incentives to demagoguery.)

Of course, Social Security has not emerged from its bouts of financial distress entirely unscathed. The 1983 Social Security legislation provided for sharply higher payroll taxation, the tapping of general revenues, and the mandatory inclusion of more individuals under the provisions of the system. It also cut benefits from what they otherwise would have been by a six-month delay of inflation indexing, the taxation of some Social Security benefits, and the gradual phasing in of a higher retirement age. In addition, enactment of legislation enabling individuals to establish individual retirement accounts (IRAs) should be viewed as a reduction in the scope that Social Security otherwise would have had. Because these measures provided for the solvency of the system in the short term, they cannot in any way be considered a rethinking of the fundamental premises of Social Security. Tinkering with the tax-benefit structure was accompanied by the vociferous expression of a commitment to the continuation of Social Security as a pay-as-you-go intergenerational transfer scheme. The retuning of specific features does not challenge the claim that the Social Security system as such remains politically untouchable.

The explanation presented here of why Social Security is, at least at present, untouchable does not rest on a claim that most net losers from Social Security reap large expressive returns from supporting it. Even a relatively small number of net losers reaping such returns can, under certain realistic conditions, translate into virtual unanimity at the polls.

Consider an electorate consisting of net losers L and net gainers G. The losers are further subdivided into L_e, those for whom expressive returns are substantial enough to motivate a vote for Social Security; and L_r, those for whom expressive returns are initially insufficient to prompt an affirmative vote. It is assumed that all those in G will vote affirmatively, whether because of expressive factors or for some other reason.

If, in an initial voting test, $G + L_e > L_r$, then a majority is seen to vote for Social Security. Once the existence of such a majority is demonstrated, that fact will affect the calculations of each of the L_r. They will recognize that p, the probability that one's vote against Social Security will be decisive, has decreased, and with it R_i. The diminished R_i will induce those of the L_r who had been near the margin to shift their vote, thus increasing the number of L_e. That shift will, in turn, further reduce R_i for the remaining L_r, and so on, until everyone who receives an expressive return, no matter how slight, from a vote for Social Security will be in the L_e camp. Although the initial effective majority in favor of Social Security may be slight, the equilibrium result is virtual unanimity.

VI. The Possibility of Fundamental Change

It has been claimed that expressive factors affect individual preferences, that the significance of expressive factors is magnified by majority rule institutions with large electorates, and that expressive returns are unusually high in the case of Social Security. If these assumptions are correct, then the model presented in this paper provides an explanation for the otherwise puzzling fact that there currently exists no politically meaningful opposition to Social Security. It also provides a basis for predicting under what conditions significant opposition could emerge.

Increased payroll taxation to ensure the continuation of benefits at the current level will, by itself, have little effect on support for the system. Higher taxation will slightly increase the return expected from a vote against Social Security, but because the probability that one's vote will be decisive is so minute, the dominant factor for most voters will continue to be the expressive returns of an affirmative vote. Survey results already reveal that young voters are skeptical of promises that they will receive benefits upon reaching retirement age; survey results also indicate, however, that young voters strongly support the continuation of the current system. It has been argued that these results are not paradoxical. Rational younger voters can be expected to persist in their support of the system even as their losses grow. Indeed, they have done so.

If, however, expressive returns were to decrease, the prevailing calculus might be altered significantly. An example may be found in principles concerning distributive justice, which are among those that can provide expressive returns. The voter who perceives himself as endorsing an equitable allocation of resources through his vote thereby receives a direct payoff. If individuals came to believe that Social Security in its present form is unfair, that it exacts burdens unmatched by compensating benefits or is in violation of individual rights, then expressive support for the system could crumble. Individuals would have an incentive to reap expressive returns by casting a ballot against practices perceived to be unfair—and politicians would have an incentive to mobilize a coalition promoting fairness in the distribution of wealth.

Increased attention to the long-term financial precariousness of Social Security can contribute to such an apprehension. One element of a moral indictment of the system is the fact that younger taxpayers are unlikely to receive retirement benefits commensurate with those of current retirees. Many studies to that effect have been made, but their political impact thus far has been minimal. What must be appended to them in order to crystallize opposition is a persuasive argument that this disparity in returns constitutes objectionable *unfairness*. It must be shown that the disparity is not due merely to an actuarial miscalculation, an unpredictable shift in demographic patterns, a temporary downturn in the economy, or some other impersonal event for which no one is to blame. One would not expect outpourings of expressive activity in response to revised actuarial projections, although the identification of actual culprits could be politically portentous. Accordingly, one can predict that if an argument charging Social Security with being a version of the chain-letter swindle gains currency, and if special interests are represented as cynically conniving to bring about unjust transfers, then substantial expressive voting against Social Security will be generated. As economists, political economists, and moral philosophers point out the true nature of the Social Security system and younger workers more fully recognize the prospect of negative returns, confidence in the system may dissolve. Ideas *do* have consequences, though not necessarily in the short run.

It is not only the young who would be induced to respond in this manner. If net gainers from Social Security become convinced that they are the beneficiaries of an unjust system, they too will be in a position to realize expressive returns by voting against the continued existence of Social Security or by voting for its radical modification. In terms simply of their own self-interest, it would be rational for each to vote to abolish a system that no one of them, were his the decisive vote, would choose to eliminate.

Expressive returns are not confined to votes motivated by moral principles. The characterization of voting behavior presented in this paper is not that it is intrinsically more moral than private activity, but that it is less consequential. Altruism is less costly in the voting booth than in the private arena—and so too is malice. An employer who displays contempt for blacks by refusing to hire them forgoes the economic benefits that their employment would bring; one who assaults an enemy risks retaliation and legal penalties. But to *vote* for policies that disadvantage blacks or one's enemies avoids these drawbacks.

If younger voters perceive themselves to be the dupes of a social policy that favors the old, they may come in increasing numbers to feel resentment or envy toward those profiting at their expense. The expression of such resentment is a consumption good, but indulging it in the private arena comes at a price: resentful behavior is generally cause for disapproval, especially if directed at a respected individual or group. Older persons will respond in damaging ways, and relations with older persons to whom one has special ties will become strained. On efficiency grounds, then, one would expect the expression of resentment to be channeled largely into political activity.

This account of the expression of malice in the voting booth bears affinity to sociological analyses of mob behavior. Both those who vote and those who are caught up in a mob are anonymous members of an aggregate within which their own contributions are unlikely to have any discernible impact on the outcome that emerges. Spared that responsibility, both can afford to indulge impulses that would, in other contexts, remain dormant. Most important for the purposes of this paper, in both cases individuals have strong incentives to do things for which they subsequently will feel regret. One who votes from malice may find himself voting for measures harsher than those he hopes will ultimately emerge; but they will indeed emerge if the strategy of expressing malice through one's vote is dominant for a large sector of the electorate. Mobs and majoritarian voting can both generate results that no one genuinely intends.

It may be unfortunate, as some analysts have alleged, that the Social Security system as presently constituted enjoys overwhelming political support. Such support means that welfare-enhancing modifications of Social Security's basic structure are now unattainable. It does not follow, however, that a substantial shift in public sentiment will produce an alternative program superior on welfare grounds. If the shift is associated with an increase in malice-inspired voting, the results may be distinctly inferior to present arrangements. It is, perhaps, excessively pessimistic to fasten on that possibility as the one most likely to be realized, but it is excessively optimistic to dismiss that prospect out of hand. What can be concluded with a high degree of confidence is the following: Social Security is an issue for which expressive factors are crucial, and those factors will remain crucial even if that which is being expressed by most voters changes. It is therefore ill-advised to suppose that policy determinations 5, 10, or 20 years hence will reflect anything closely resembling standard economic notions of an efficient outcome.

References

Beck, Nathanial. "A Note on the Probability of a Tied Election." *Public Choice* 18 (Fall 1975): 75–80.

Brennan, Geoffrey, and Buchanan, James. "Voter Choice and the Evaluation of Political Alternatives." Center for Study of Public Choice, 1982. Mimeo.

Brennan, Geoffrey, and Lomasky, Loren. "Large Numbers, Small Costs: The Uneasy Foundations of Democratic Order." Center for Study of Public Choice, 1982. Mimeo.

Brennan, Geoffrey, and Lomasky, Loren. "Institutional Aspects of 'Merit Goods' Analysis." *Finanzarchiv* 41 (1983): 183–206.

Browning, Edgar K. "Why the Social Security Insurance Budget Is Too Large in a Democracy." *Economic Inquiry* 13 (September 1975): 373–88.

Browning, Edgar K. "The Politics of Social Security Reform." In *Financing Social Security Reform*, pp. 187–207. Edited by Colin D. Campbell. Washington, D.C.: American Enterprise Institute, 1979.

Buchanan, James, and Tullock, Gordon. *The Calculus of Consent*. Ann Arbor: University of Michigan Press, 1962.

Downs, Anthony. *An Economic Theory of Democracy*. New York: Harper & Row, 1957.

Edelman, Murray. *The Symbolic Uses of Politics*. Urbana: University of Illinois Press, 1964.

Ferejohn, John, and Fiorina, Morris P. "The Paradox of Not Voting: A Decision Theoretic Approach." *American Political Science Review* 68 (1974): 525–36.

Ferrara, Peter J. *Social Security: The Inherent Contradiction*. Washington, D.C.: Cato Institute, 1980.

Goodfellow, Gordon, and Pellechio, Anthony. "Individual Gains and Losses from Social Security before and after the 1983 Amendment." *Cato Journal* 3 (Fall 1983): 417–42.

Hartwig, Richard. "Rationality and Voting." *Journal of Political Science* 9 (1981): 1–6.

Kiewet, D. R., and Kinder, D. R. "Economic Discontent and Political Behavior: The Role of Personal and Collective Economics Judgments in Congressional Voting." *American Journal of Political Science* 23 (1979): 495–527.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. 2d ed. Chicago: University of Chicago Press, 1970.

Kuhn, Thomas S. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press, 1977.

Laudan, Larry. *Progress and Its Problems: Toward a Theory of Scientific Growth*. Berkeley: University of California Press, 1977.

Mitchell, William C. *The Popularity of Social Security: A Paradox in Public Choice*. Washington, D.C.: American Enterprise Institute, 1977.

Munro, Douglas R., and Parsons, Donald O. "Intergenerational Transfers in Social Security." In *The Crisis in Social Security*, pp. 65–86. San Francisco: Institute for Contemporary Studies, 1977.

O'Neill, June A. "Alternatives for Financing a Mature Social Security System." In *Financing Social Security Reform*, pp. 173–85. Washington, D.C.: American Enterprise Institute, 1979.

Ordeshook, P., and Riker, W. *An Introduction to Positive Political Theory*. Englewood Cliffs, N.J.: Prentice Hall, 1973.

Patton, Carl V. "The Politics of Social Security." In *The Crisis in Social Security*, pp. 147–71. Edited by Michael J. Boskin. San Francisco: Institute for Contemporary Studies, 1977.

Quine, W. V. O., and Ullian, J. L. *The Web of Belief*. 2d ed. New York: Random House, 1978.

Report of the National Commission on Social Security Reform. Washington, D.C.: Government Printing Office, January 1983.

Stigler, George J. "General Economic Conditions and National Elections." *American Economic Review* 63 (May 1973): 160–67.

Tufte, E. R. *Political Control of the Economy*. Princeton: Princeton University Press, 1978.

Tullock, Gordon. *Toward a Mathematics of Politics.* Ann Arbor: University of Michigan Press, 1968.

Tullock, Gordon. "The Charity of the Uncharitable." *Western Economic Journal* 9 (December 1971): 379–92.

Weaver, Carolyn. "The Long-Term Outlook for Social Security—Continued Political Turmoil." Washington, D.C.: National Commission on Social Security Reform, October 1982. Mimeo.

THE SAFETY AND EFFICACY OF NEW DRUG APPROVAL

Dale H. Gieringer

> The perception that agencies are out of control arises from the fact that in being called on to make fundamental value judgments they have moved outside their accustomed sphere of activity, outside their expertise, and outside the established system of controls. This perturbation of the regulatory process will not be corrected until the agencies are relieved of the necessity of making judgments they are not equipped to make.
>
> —Richard M. Cooper[1]

Introduction

The U.S. Food and Drug Administration (FDA) has been widely criticized for obstructing the development and use of new pharmaceutical drugs. Under present law, no new drug or medical device may be brought to market until it has been officially approved as "safe and effective" by the FDA. Unapproved products may not legally be sold and are available only on a highly restricted basis in investigational studies specifically preapproved by the FDA. Critics have frequently accused FDA of failing to permit potentially valuable drugs on the market. The ironic effect is that consumers are denied access to drugs that could benefit them on alleged grounds of public health.

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is a policy consultant with the Decisions and Ethics Center of the Department of Engineering-Economic Systems at Stanford University.

[1] Richard M. Cooper (FDA chief counsel), "The Role of Regulatory Agencies in Risk-Benefit Decision-Making," *Food, Drug and Cosmetic Law Journal* 33 (December 1978): 772.

The present approval system was established by the Food, Drug and Cosmetics Act of 1938, which required that all new drugs be approved for safety by the FDA; this was extended to include medical devices in 1976. Approval criteria were substantially strengthened in 1962 by the Kefauver-Harris amendments, which added the requirement that drugs be proven "effective" as well as safe. In addition, the 1962 amendments placed stringent controls on the use of investigational drugs, which had previously not been regulated.

In the wake of the 1962 amendments, the FDA's regulation of new drugs became increasingly stringent and risk averse, leading critics to complain of overregulation. The bureaucracy at the FDA acquired a reputation for remarkable inefficiency and delay, with processing times for new drug applications (NDAs) typically extending to two years or more, and paperwork reaching into tens of thousands of pages.[2] Meanwhile, the time and expense of new drug development increased dramatically. By 1976 the cost of developing a new drug had risen to an estimated $24 million, 10 to 20 times as much as in the early 1960s, while development times had climbed from a couple of years to the better part of a decade.[3]

By the mid-1970s critics began to argue that FDA regulation was causing the United States to suffer a so-called drug lag, a slowdown in the development of new pharmaceutical drugs. Many critics pointed to a dramatic decline in the number of innovative new drugs approved by the FDA: from an average of 50 per year in 1955–60 to only 17 per year in 1965–70 and after.[4] In a notable cost-benefit study of the 1962 amendments, Sam Peltzman argued that the consumer costs of the post–1962 slowdown in drug development substantially outweighed benefits from improved drug efficacy and safety.[5] It is unclear, however, how much of the apparent slowdown in new drug development can be attributed to FDA regulation since it began some months before the 1962 amendments took effect, and a similar decline occurred in foreign countries.[6]

---

[2] U.S. General Accounting Office (GAO), *FDA Drug Approval—A Lengthy Process that Delays the Availability of Important New Drugs,* Report to the House Committee on Science and Technology, Subcommittee on Science, Research and Technology (May 1980).

[3] W. Wardell, M. Hassar, S. Anavekar, and L. Lasagna. "The Rate of Development of New Drugs in the United States, 1963–1975," reprinted in U.S. House of Representatives, *The FDA's Process for Approving New Drugs: Hearings of the House Committee on Science and Technology, Subcommittee on Science, Research and Technology* (96th Congress, 1979), pp. 543–63.

[4] Henry G. Grabowski and John M. Vernon, *The Regulation of Pharmaceuticals: Balancing the Benefits and Risks* (Washington, D.C.: American Enterprise Institute, 1983), pp. 29–30.

[5] Sam Peltzman, *Regulation of Pharmaceutical Innovation: The 1962 Amendments* (Washington, D.C.: American Enterprise Institute, 1974).

More substantive evidence for drug lag is that by the late 1960s new drugs were coming into use in foreign countries months and years before receiving FDA approval. In a comparative study of the United States and Great Britain, William Wardell and Louis Lasagna found that U.S. physicians were at a relative disadvantage in the number of potentially valuable new drugs available to them because of the greater stringency of FDA regulation.[7] A General Accounting Office (GAO) study found that in four out of five foreign countries, regulatory approval times were 6 to 18 months shorter than in the United States.[8] Meanwhile, there was evidence of a growing number of so-called orphan drugs—potentially useful drugs that were not being brought to the U.S. market because the costs of FDA approval would exceed their potential sales revenues.[9] This evidence was accompanied by reports of patients suffering from lack of access to unapproved drugs or denial of FDA permission for investigational treatment; other Americans sought treatment abroad or resorted to black- or gray-market suppliers for unapproved products.[10]

Until recently, the problem of FDA overregulation received little public attention. Throughout the late 1960s and 1970s critics complained that the political pressures on the FDA were such as to systematically favor overregulation. This complaint was explained by the fact that new drug accidents, such as with thalidomide, were more widely publicized by the press than casualties from the drug lag, whose victims were largely anonymous.[11] However, criticism of the drug-lag problem finally mounted to the point where the FDA was forced to respond. In 1979 Congress held hearings on the drug-lag problem,[12] prompting the FDA to establish a new "fast-track" approval system for important new drugs. Reforms to streamline new drug regulation were further pressed by the Reagan administration. These efforts succeeded in reducing approval times for important new drugs by as much as 40 percent, to an average of 19 months, while new drug approvals reached their highest level since the early 1960s.[13]

---

[6] Donald Kennedy, "A Calm Look at 'Drug Lag'," *Journal of the American Medical Association* 239 (1978): 423–26.

[7] William Wardell and Louis Lasagna, *Regulation and Drug Development* (Washington, D.C.: American Enterprise Institute, 1974).

[8] GAO, "FDA Drug Approval," p. 7.

[9] Louis Lasagna, "Who Will Adopt the Orphan Drugs?" *Regulation* (November/December 1979): 27–32.

[10] For examples, see John Kelly, "Bridging America's Drug Gap," *New York Times Magazine*, 13 September 1981; the testimonies of William Regelson and J. Kiffin Penry, in *The FDA's Process*, pp. 288–89 and p. 101; and David L. Shanks, "Chance Denied," letter to the *Wall Street Journal*, 4 April 1985, p. 33.

[11] Compare Milton Friedman, *Free to Choose* (New York: Avon Books, 1980), pp. 193–200.

[12] *The FDA's Process.*

While the thrust of recent FDA reforms has been to speed new drug approvals, the drawback of this strategy is that it inevitably raises the risk of new drug accidents. Efforts at deregulation have accordingly been attacked by consumerist opponents on grounds of safety. These concerns were born out in 1982 by the highly publicized case of Oraflex, an innovative arthritis drug that received "fast track" approval from the FDA—partially on the basis of foreign data, which had hitherto been disallowed by FDA policy. Oraflex subsequently had to be withdrawn from the market when it proved toxic to 61 British and 11 American victims, prompting critics to call for stricter regulation. Some patients, however, claimed to enjoy unique benefits from Oraflex, which may have had the unique ability to arrest progress of arthritis.[14]

The Oraflex dispute illustrates a fundamental problem of the present approval system for new drugs; namely, the assumption that new drugs be approved as "safe and effective" on a collective, societywide basis. The problem is that safety and efficacy are inherently subjective concepts, whose meaning inevitably varies from individual to individual. As noted by one observer, "No one has yet defined safety and efficacy. Nevertheless, distinguished panels attempt to make what are termed 'scientific assessments' in the absence of objective basing points."[15] In practice, safety and efficacy depend strongly on individual circumstances such as age, sex, genetic makeup, and a host of other medical and personal factors that are often difficult for regulators to know. Even more important, the meaning of safety and efficacy depends crucially on personal values and attitudes toward risk: what seems safe to one person may well seem unsafe to another in similar circumstances. By imposing collective choice in drug risk, the present system is therefore inherently controversial, requiring the arbitrary imposition of values by technocratic authority. In this light, the debate over drug lag can be understood as a value dispute between "pharmacophobes" and "pharmacophiles," that is, those who are respectively more and less risk averse to new drug hazards.

---

[13] The FDA has also recently announced further reforms to reduce new drug approval times to 17 months: *Federal Register*, 22 February 1985, p. 7452.

[14] *Wall Street Journal*, 3 August 1982, p. 3; *WSJ* 24 November 1982, p. 12; *WSJ* 27 December 1982, p. 5.

[15] Joseph Cooper, "Purpose, Technique and Strategy in the Regulation of New Drugs," in Richard Landau, ed., *Regulating New Drugs* (Chicago: University of Chicago Press, 1973), p. 30.

An evident solution to the drug safety problem is to make all drugs available for those who want them with appropriate informational warnings. In principle, this might be done through a system of "informed choice," where the basic object of drug policy would be to inform, not to restrict, consumer choice. Under such a policy, the basic role of the FDA would no longer be to determine whether or when drugs could be made available, but rather to assure that patients were adequately informed of the risks. In contrast, the present regulatory system often fails to provide consumer information on drug hazards. For example, prescription drugs are regularly sold without consumer labeling or warnings from the manufacturer. Thus Oraflex patients had no way of knowing that it was a relatively untested drug when it first came on the market unless their physicians or pharmacists happened to tell them. Under an informed-choice policy, new drugs like Oraflex would be sold with appropriate warnings about the increased risk of unknown new reactions, so that patients and physicians could decide for themselves whether to try them. In effect, the drug-lag debate would be resolved through consumer choice in the market.

Informed choice has occasionally been mentioned as an attractive alternative to the present system of mandatory approval,[16] but rarely seriously discussed.[17] Informed choice has been virtually ignored in legislative reform proposals, as the present regulatory system has developed with little consideration of other alternatives. In the meantime, the FDA has in some ways actively discouraged informed choice, for example, by exempting manufacturers from providing written consumer warnings with prescription drugs.[18] As with most FDA regulations, the consequences of this policy have never been fully evaluated.[19]

---

[16] Henry Grabowski, *Drug Regulation and Innovation* (Washington, D.C.: American Enterprise Institute, 1976), p. 82; and Grabowski and Vernon, "Regulation of Pharmaceuticals," pp. 71–72; compare Kenneth Arrow, "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review* 53 (1963): 967.

[17] For a full analysis of informed choice in drugs, see Dale Gieringer, "Consumer Choice and FDA Drug Regulation," Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University (1984).

[18] See Peter Temin, "The Origin of Compulsory Drug Prescriptions," *Journal of Law and Economics* 22 (April 1979): 91–105.

[19] For further analysis of the present prescription drug system, see Peter Temin, *Taking Your Medicine: Drug Regulation in the United States* (Cambridge, Mass.: Harvard University Press, 1980).

This paper considers the costs of FDA regulation of new drugs in greater detail. The following section, in accordance with Peltzman's results, argues that the costs of regulatory approval may well outweigh the benefits. But, unlike Peltzman's analysis, this argument is based entirely on an examination of the real public health benefits and risks of available drugs. More important, it will be shown that new drug casualties could be reduced at least as effectively through informational warnings as through regulatory prohibition from the marketplace. From this analysis, it follows that an informed-choice policy would be superior to any new-drug approval system, no matter how well balanced the latter system is from a cost-benefit viewpoint. This conclusion stems from the fact that the informed-choice policy would not require sacrificing drug benefits for one class of patients in order to protect others.

Risk-Benefit Tradeoffs of FDA Approval

The FDA typically has avoided any attempt to apply cost-benefit methodology to the regulation of new drugs, and it has tended systematically to overstate the dangers of approving new drugs relative to the dangers of delaying them.[20] There has been one cost-benefit study with conclusions apparently favorable to the FDA—namely, that of James Jondrow, who argued that the consumer benefits of the 1962 efficacy requirement exceeded the expenses of additional testing. Jondrow's study, however, is flawed by its failure to account for the cost of delay in introducing new drugs.[21] This shortcoming was addressed in Peltzman's cost-benefit study. Based on an econometric analysis of drug company sales, Peltzman estimated the cost to consumers of the 1962 amendments at $350–450 million versus benefits of only $100 million.[22] The assumptions of Peltzman's study, however, have been disputed on various grounds, including charges of there being certain problems of economic methodology.[23]

---

[20] Rita Ricardo-Campbell, "Risk-Benefit, Cost-Benefit: Improving Government Regulation of Approval of New Drugs," in J. Van Der Gaag, W. B. Neenan, and T. Tsukahara, Jr., eds., *Economics of Health Care* (New York: Praeger Publishers, 1982), chap. 2.

[21] James Jondrow, "A Measure of the Monetary Benefits and Costs to Consumers of the Regulation of Prescription Drug Effectiveness," Ph.D. dissertation, University of Wisconsin (1972); as discussed in Leonard Schifrin, "Lessons from the Drug Lag," report prepared for the Office of Technology Assessment (June 1980). For further discussion of FDA cost-benefit studies, see Temin, *Taking Your Medicine*, pp. 141–51; and Grabowski and Vernon, pp. 37ff.

[22] Peltzman also estimated an additional $200 million in unmeasurable costs owing to delays in approving particularly efficacious drugs, versus less than $50 million in unmeasurable benefits from reduced new drug casualties. Peltzman, p. 81.

Perhaps the most important criticism of Peltzman's study is that it assumes that the entire post–1962 slowdown in new drugs is attributable to regulation. Yet, as noted above, there is considerable evidence to suggest that the slowdown was at least partially independent of the 1962 amendments. In fact, a good case can be made that it was at least partly a voluntary response by the drug industry to increased public demand for drug safety in the wake of the thalidomide tragedy.[24] Nevertheless, economic studies have generally concurred that FDA regulation has had some adverse effects on new drug development.[25] For example, in a more recent study relying exclusively on data from the 1970s, Steven Wiggins has argued that regulation reduced new drug introduction rates by 60 percent.[26]

A second criticism of Peltzman's study is that his economic methodology evaluates drug costs and benefits exclusively in monetary terms, and in particular on the basis of revenues earned by drug manufacturers. The connection between industry revenues and real health benefits, however, seems dubious at best, given that the market for worthless and even harmful patent medicines has often been extremely lucrative.[27] In addition, the attempt to evaluate human life and health in monetary terms involves controversial problems, since it is not always possible to place a specific monetary value on life.[28]

Because of these criticisms, it is useful to try to evaluate the costs and benefits of regulation in terms of real public health effects. In

[23] For critiques of Peltzman's study, see Richard Nelson and Thomas Spavins, "An Evaluation of Consumer Protection Legislation: The 1962 Amendments, A Comment," *Journal of Political Economy* 83 (1975): 655–61; and testimony of Leonard Schifrin and Samuel Baker in U.S. Senate, Select Committee on Small Business, *Hearings on Competitive Problems in the Drug Industry* (93rd Cong., 1st sess., 1973), Part 23, pp. 9766–9801.

[24] Temin, *Taking Your Medicine,* pp. 146–48.

[25] For additional discussion of FDA cost-benefit studies, see Grabowski and Vernon, pp. 37ff.; and Temin, *Taking Your Medicine,* pp. 141–51.

[26] Steven Neil Wiggins, "Product Quality Regulation and New Drug Introductions: Some New Evidence from the 1970's," *Review of Economic Statistics* 63 (November 1981): 615–19; and Wiggins, "The Impact of Regulation on Pharmaceutical Research Expenditures: A Dynamic Approach," *Economic Inquiry* 21 (January 1983): 115–28.

[27] See James Harvey Young's account of the early U.S. patent medicine market, *The Toadstool Millionaires* (Princeton, N.J.: Princeton University Press, 1961).

[28] While it is often possible to impute a monetary value to life on the basis of consumers' willingness to pay for risk reduction, this is generally true only at limited levels of risk. At higher levels of risk, the monetary value of life tends to increase, becoming infinite beyond a certain point. For an analysis, see Ronald Howard, "On Making Life and Death Decisions," in R. C. Schwing and W. A. Albers, Jr., eds., *Societal Risk Assessment* (Milford, Mich.: General Motors Research Labs, 1980).

doing so, it should be noted that any estimate of drug costs and benefits is inevitably fraught with uncertainties. It is generally not possible to know what would have happened in the absence of regulation, especially with regard to any additional drug casualties that may have occurred or with regard to other new drugs that may have been developed. In addition, there is a notable lack of reliable information on the actual health benefits and costs of existing medicine. A simple, rough estimate of the real health costs and benefits of drug regulation, however, can be obtained by considering the two most salient, relevant data sets presently available. One set consists of known casualties from new drug reactions in the United States and foreign countries, as reported in the world press and medical journals. The other set consists of reductions in the mortality rate, as reported by the U.S. Bureau of the Census, for diseases in which it is generally agreed that advances in drug therapy have played a major role in recent decades.

With regard to known casualties, it is possible to use worldwide new drug reaction reports to obtain a rough gauge of the value of FDA new drug approval. This is because most new drug casualties have occurred in foreign countries where regulation has been more liberal than the United States, at least until recent years. The difference between U.S. and foreign casualty rates may thus be taken as a rough gauge of the marginal benefits of the post–1962 FDA regime.

With regard to mortality rate declines, the benefits of new drugs may be estimated from the reported declines in mortality rates for major diseases recorded in U.S. vital statistics. (Unfortunately, corresponding statistics on disease morbidity are lacking, so it is necessary here to confine the discussion of new drug benefits entirely to the prevention of deaths.) By estimating the contribution of new drugs to observed mortality rate declines, it is possible to estimate the potential cost of delay in new drug approval. This may be done by assuming that the effect of delay is to postpone all subsequent new drug benefits accordingly. Thus, if a new drug saves 1,000 lives per year, a one-year delay in its introduction may be estimated to cost 1,000 lives. While this is clearly a simplifying assumption, which is subject to qualification in many instances, it may serve as a useful rough indication of the potential costs of new drug delay.

Cost of Delay in New Drug Introductions

Table 1 contains a summary of mortality rate declines for diseases where advances in drug therapy are thought to have played a signif-

TABLE 1

ESTIMATED MORTALITY RATE REDUCTION ATTRIBUTABLE TO NEW DRUGS: 1950–77

	Mortality per 100,000 Age Adjusted (1977 Base)			Est. Reduction in Mortality Rate Attrib. to Drugs	
Disease	1950	1960	1977	1950–60	1960–77
Hypertension[a]	74.8	53.8	24.0	6–12	10–21
Cerebrovascular Disease	142.0	136.0	84.0	3–6	20–33
Tuberculosis[b]	23.4	6.5	1.4	6–10	1–4
Pneumonia	32.6	37.1	23.1	—	2–6
Kidney Infection	2.3	4.9	1.7	—	0.3–2
Polio	1.3	0.0	0.0	1.3	—
Rheumatic Fever	16.4	11.0	5.9	0–1	0–1
Nephritis and Nephrosis	21.0	8.8	3.9	1–3	—
Meningitis	1.0	1.1	0.7	0–1	0–1
Peptic Ulcers	6.1	7.1	2.7		
Asthma	3.3	3.2	0.8		
TOTAL REDUCTION IN MORTALITY RATE[c]				17–34	33–68

[a] Data for hypertension and hypertensive heart/renal disease are adjusted to their 1950 base definition using adjustment factors supplied in U.S. Bureau of the Census, "Comparability of Mortality Statistics for 7th and 8th Revisions of the International Classification of Diseases," *U.S. Vital and Health Statistics*, Series 2, No. 66. (Washington, D.C.: Govt. Print. Office, various years).

[b] Drugs to treat tuberculosis were introduced in 1946–47; see also accompanying text.

[c] Estimates refer to average annual reductions in mortality rates per 100,000 people.

SOURCE: Cols. 1–3, U.S. Bureau of the Census, National Center for Health Statistics, *Vital Statistics of the United States*, various years. Cols. 4–5, author's estimates based on mortality statistics and discussions with medical experts (supra, note 30).

icant role in the period 1950–77.[29] In the two righthand columns is displayed, for each disease, a range of plausible estimates of that portion of reduced mortality that may reasonably be attributed to drugs. Unfortunately, there is a lack of epidemiological and drug usage statistics to directly measure the total impact of drug therapy in the population (with the important exception of hypertension and stroke, as described below). Therefore, the estimates in Table 1 are by necessity "best guess" estimates, based on the author's reading of mortality statistics and discussions with medical experts.[30]

---

[29] For a more detailed discussion see Gieringer, "Consumer Choice."

Estimates of the contribution of drugs in Table 1 have been adjusted to discount preexisting secular trends in mortality, and have been further reduced on the side of conservatism where considerable uncertainty or disagreement among experts is apparent or where a prominent contribution from other extraneous factors seems likely. The estimates in Table 1 are for the period 1950–77, and have been adjusted to exclude the impact of any drugs introduced before 1950. While mortality statistics have been broken down into two periods, 1950 to 1960 and 1960 to 1977, no attempt is made to separate the effects of drugs introduced before and after 1960.

There is little question that the greatest mortality savings attributable to drugs since 1950 have occurred in the related areas of hypertension and cerebrovascular disease or stroke. Mortality from hypertension has declined in parallel with the introduction of antihypertensive drugs, beginning in the 1950s. This has been accompanied by a parallel decline in stroke, for which hypertension is a leading risk factor.[31] It is widely believed that drugs have been a major factor in the control of hypertension and stroke. Most experts estimated that drug therapy has reduced mortality rates by 60–75 percent.[32]

Fortunately, there exist drug-use surveys from which it is possible to confirm this estimate. The National Health Examination Survey of 1960–62, for example, found that 11.9 percent of all hypertensives were under therapeutic control. Following a nationwide public education campaign in the early 1970s, control levels of 45 percent or more were being reported.[33] All in all, it may reasonably be estimated that 35–50 percent of the hypertensive population came under control in the period 1950–76, or 25–40 percent since 1960. Given that about 25 percent of the adult population is hypertensive, that this proportion accounts for 85 percent of all cases of stroke, and that antihypertensive drugs have been shown to reduce the risk of stroke by a factor of 3 or 4,[34] it can be estimated that stroke mortality would be reduced by some 60 percent if all hypertensives were treated. From the drug-use surveys, therefore, it follows that drugs would have caused total stroke mortality rates to decline by 21–30 percent from 1950 to 1976, or by 15–24 percent since 1960—declines that are in agreement with the estimates given in Table 1.

---

[30] Various expert physicians and pharmacologists were asked to help estimate the percentage of disease mortality declines attributable to drugs as precisely as possible. In many cases it was felt that exact numbers could not be meaningfully assigned, and many answers were qualified to exclude the effect of secular trends or other uncertain factors. While there was an unavoidably conjectural element in this process, numerical estimates tended to converge in precisely those areas where it was felt that the contribution of drugs had been greatest, whereas greater divergence of opinion was expressed where the contribution of drugs seemed more marginal.

[31] Hypertension has been associated with 85 percent of all cerebrovascular problems. See W. B. Kannel, P. A. Wolf, J. Verter, and P. M. McNamara, "Epidemiological Assessment of the Role of Blood Pressure in Stroke: The Framingham Study," *Journal of the American Medical Association* 214 (1970): 301ff.

[32] Compare Boranhi's estimate that "at least 40 percent" of the observed decline for hypertension is attributable to better treatment. See N. O. Boranhi, "Mortality Trends in Hypertension," in National Heart, Blood, and Lung Institute, *Proceedings of the Conference on the Decline in Heart Disease Mortality*, NIH Publication 79-1610 (October 1978), p. 227.

A major breakthrough in tuberculosis therapy occurred with the introduction of streptomycin and PAS in 1947. Although these drugs were later displaced by others, the major advances must be said to have begun three years before the start of the period covered here. For this reason, mortality savings for tuberculosis as included in Table 1 should be considered optional. Also, in estimating the contribution of drugs to tuberculosis mortality savings, it is necessary to adjust for a preexisting, secular mortality decline of over 4 percent per year dating back to the early 1900s.[35] Discounting for this, about one-half of the observed post–1950 mortality decline seems attributable to drugs.

Antibiotics account for most of the other gains shown in Table 1. It is generally agreed that antibiotics have played a major role in the control of pneumonia and kidney infections plus some other conditions for which separate mortality data are lacking. They also appear to have played a more modest and uncertain role in the treatment of other diseases, such as meningitis. In the case of pneumonia and other infections, it is necessary to adjust for occasional upswings in mortality statistics owing to epidemics, although the long-term trend has been downward.[36] In addition, it is necessary to adjust for preexisting mortality declines owing to the use of penicillin, which dates from around 1941 (well before the 1950 start date used here). Nevertheless, new antibiotics have been necessary to deal with the resistant strains of bacteria that are continually emerging, as well as to treat patients with special allergies.

---

[33] For a summary of evidence, see G. W. Ward, "Changing Trends in Control of Hypertension," *Public Health Reports* 93 (January/February 1978): 31–34; M. P. Stern, "The Recent Decline in Ischemic Heart Disease Mortality," *Annals of Internal Medicine* 91 (1979): 630–40; and Boranhi, "Mortality Trends."

[34] H. M. Perry, "The Treatment of Mild Hypertension," in Elliot Rapaport, ed., *Cardiology Update: Reviews for Physicians* (New York: Elsevier Publishing, 1981), pp. 145–63.

[35] For a similar analysis, see Peltzman, pp. 58–63.

[36] Thus, the anomalous rise in pneumonia deaths around 1960 was caused by the Asian flu epidemic. It seems likely that flu vaccines have been a factor in the subsequent decline of pneumonia mortality to historic lows.

Other areas in which advances in drugs have been cited include ulcers, asthma, and kidney conditions such as nephrosis.[37] However, in most of these areas, the impact of drugs seems marginal or uncertain at best. One exception is polio, which has been largely eradicated as a result of the widespread use of polio vaccines.

Not included in Table 1 are advances in cancer chemotherapy, for which the National Cancer Institute attributes some 11,000 to 46,000 cures per year.[38] These are omitted here because of a lack of adequate cancer mortality statistics and because most of the cures appear to be attributable less to the introduction of new drugs than to the more effective application of existing ones.[39]

In sum, Table 1 shows estimated mortality reductions attributable to the use of new drugs of 50–102 per 100,000 from 1950 to 1977 (or 43–88 per 100,000 if tuberculosis is excluded). At this rate, it follows that a one-year delay in new drug benefits would cost 37,000 to 76,000 lives per decade in the U.S. population (32,000 to 65,000 excluding tuberculosis). By comparison, FDA delays in approving new drugs have often been estimated at two years or more.

Caution must be used in extrapolating from these figures to estimate new drug approval costs. The assumption that every delay in approval postpones all subsequent drug benefits equivalently may in fact result in an overestimation of the costs of delay. It seems likely that some drug benefits are partly attributable to events that are independent of the exact date of new drug approval. For example, the use of antihypertensive drugs languished for many years until finally being stimulated by public education efforts over a decade after the first drugs were approved. However, this problem may well have been aggravated by restrictions on drug labeling and advertising, which prohibit the mention of any drug uses or indications that have not been specifically approved by the FDA.[40] For many years, the FDA restricted approval of antihypertensives to cases of severe high blood pressure only and may have accordingly retarded their application to moderate hypertension, where they have also proved valuable.[41]

---

[37] The Pharmaceutical Manufacturers' Association's *Prescription Drug Industry Factbook* for 1980 (p. 46) also lists arteriosclerosis and infectious hepatitis, but with little apparent foundation.

[38] According to the director of the National Cancer Institute, *Washington Post,* 18 October 1981, p. A-15.

[39] See Alfred G. Gilman, Louis S. Goodman, and Alfred Goodman, *The Pharmacological Basis of Therapeutics,* 6th ed. (New York: Macmillan, 1980), p. 1249. Nevertheless, it has been reported that the use of certain cancer drugs has been impeded by the difficulty of obtaining NDA approval for new indications. See the testimony of C. Gordon Zubrod, in *Hearings on Competitive Problems,* Part 23, pp. 9672–88.

Another factor that may reduce the impact of FDA delay is that the delay itself may bolster consumer confidence in drug safety, thereby enabling new drugs to reach a wider market more quickly once they are finally approved. However, this factor is not relevant where delays are caused by sheer bureaucratic inefficiency. In this connection, it should be noted that approval times for important new drugs have been reported to average from 8 to 19 months in the wake of recent FDA reforms.[42] At the lower rate, the cost of regulatory delays could be estimated between 21,000 to 51,000 lives per decade; at the higher rate, it could be estimated at between 51,000 and 120,000 lives per decade.

Whatever the accuracy of these estimates, there are important respects in which they may actually understate the true costs of regulation. In particular, they completely ignore the drug benefits of reduced morbidity from crippling strokes, polio, and other nonfatal illnesses, the value of which in many cases may be comparable to that of life itself.[43] Moreover, they do not account for lost benefits from drugs that may otherwise have been marketed but were entirely suppressed by regulation during the period of mortality data covered here. In short, while numerical estimates are admittedly rough, the evidence suggests that regulatory delays in new drug approval may be quite costly, with casualties on the order of tens of thousands of lives per decade.

In confirmation of these conclusions, it is useful to cite one or two specific cases in which FDA approval delays have been plainly costly.

---

[40] Current law does not actually forbid the use of approved drugs in unapproved indications; however, only FDA-approved indications may be mentioned in labeling or usage instructions, and many doctors are reportedly reluctant to prescribe such drugs for unapproved indications.

[41] Testimony of William Wardell, in *The FDA's Process,* p. 62.

[42] "FDA Approval of New Drugs is Speedier, But More Progress is Needed, GAO Says," *Wall Street Journal,* 16 September 1981, p. 8.; "DHHS New Drug Regulations," *Federal Register,* 22 February 1985, p. 7452.

[43] In this connection, Peltzman estimated the economic benefits of tranquilizers in reduced hospital days to be greater than the total mortality and morbidity savings attributable to tuberculosis drugs and polio vaccines; however, he neglected to account for the considerable costs of addiction, accidents, and adverse reactions owing to tranquilizers. Peltzman, pp. 63–66.

Most of the mortality reductions shown in Table 1 happen to be from drugs approved prior to 1962, when FDA approval delays were relatively small. While this fact may itself possibly be attributed to the tightening of FDA regulation after 1962, it may also be attributed to extraneous factors.

However, one important class of drugs that clearly suffered from post–1962 FDA regulation was the beta blockers, an innovative treatment for a variety of cardiovascular conditions. The FDA approved the first U.S. beta blocker, propranalol, in 1968—three years after it had been approved in Great Britain. In subsequent years, the FDA was criticized for delaying the introduction of other, newer beta blockers and for restricting approval of propranalol to only limited indications, against the advice of expert cardiologists.[44] Finally, in November 1981, the FDA announced its approval of a new beta blocker, timolol, for an innovative indication, the prevention of second heart attacks. The FDA's action was based on a study published seven months earlier, showing that timolol could reduce mortality from second heart attacks by enough to save an estimated 6,500 to 10,000 lives per year in the United States.[45] At this rate, it can be estimated that some 4,000 to 5,800 preventable deaths occurred during the seven months required by the FDA for its purportedly expedited approval. However, as noted by Wardell, there had been clinical evidence for the efficacy of beta blockers in preventing second heart attacks as early as 1974.[46] The total cost of this seven-year delay could then be put at some 45,000 to 70,000 lives—several times greater than all the casualties resulting from thalidomide and other major new drug disasters.[47]

FDA regulation also may be responsible for major adverse public health effects in the area of cancer prevention. In recent years, there has been growing evidence that certain vitamins and minerals may have significant potential for blocking cancer.[48] While the evidence is far from certain, it seems possible that substantial benefits could be obtained from prophylactic use of vitamins and minerals as dietary supplements.[49] Vitamins and minerals are presently classified as "food supplements" and are exempt from drug regulation under special legislation that Congress passed in 1975 to prevent the FDA from banning over-the-counter sales of megavitamins. However, it is illegal for manufacturers to make any reference to possible health benefits of vitamins without becoming subject to new drug application (NDA) approval requirements for proof of efficacy. In prohibiting the advertising of possible anticarcinogenic benefits of vitamins and minerals, present regulations may be having a substantially adverse effect on consumer education and health. For example, assuming that food supplements could reduce the risk of cancer by 10 percent, some 3,500 lives per year could be saved if only 10 percent of the population were persuaded to take them.

---

[44] Wardell and Lasagna, pp. 61–64, 110–13.

[45] According to FDA estimates; quoted in the *Wall Street Journal*, 27 November 1981, p. 5.

[46] W. Wardell, "Are These Requirements Enough or Too Much?" in A. F. De Schaepdryver, L. Lasagna, F. H. Gross, and D. R. Laurence, eds., *The Scientific Basis of Official Regulation of Drug Research and Development*, Proceedings of the 7th International Congress of Pharmacology (1978), reprinted in *The FDA's Process*, pp. 527–42.

[47] Theoretically, given that the alternative beta blocker propranalol was available in the U.S. market, there was nothing to prevent physicians from venturing to prescribe it for second heart attacks, even though this was an unapproved and unproven indication for this drug. However, as noted previously, physicians are usually reluctant to prescribe for unapproved indications. Compare the *Wall Street Journal* editorial of 2 November 1981, "100,000 Killed," and the response by FDA Commissioner Arthur Hayes, Jr., *WSJ*, 18 November 1981 (letter to the editor).

Finally, in projecting the costs of new drug delays, the question must be asked as to whether drug development can be expected to continue at its historic rate. In recent years there have been such major developments (not shown in Table 1) as the introduction of cimetidine for ulcers, cyclosporine for transplants, and the use of beta blockers to prevent second heart attacks. As noted above, beta blockers have been estimated to have the potential of saving 6,500 to 10,000 lives per year, and comparable gains may be achieved through the use of various other kinds of heart drugs now being developed.[50] Even greater gains may result from advances in cancer treatment through recombinant DNA technology; a mere 10 percent reduction in cancer mortality would save 35,000 lives per year.

Costs of Reactions to New Drugs

Table 2 lists the casualties associated with the introduction of new drugs. The table is limited to incidents in which more than 100 deaths

---

[48] Vitamins A, C, and selenium have commonly been mentioned. For further discussion, see Bruce Ames, "Dietary Carcinogens and Anticarcinogens," *Science* 221 (23 September 1983): 1256–64.

[49] There is evidence that these effects could range as high as 50 percent in the case of certain cancers. Although many authorities, including a recent panel of the National Research Council of the National Academy of Sciences, have begun to advocate dietary changes to prevent cancer, most have shied away from advocating the use of supplements for lack of further evidence. See, for example, "Scientists Conduct Research on Nutrients That May Block Cancer," *Wall Street Journal*, 15 November 1982, p. 1.

[50] "Anti-cholesterol Treatment Can Cut Risk of Heart Attacks Up To 50%, Study Says," *Wall Street Journal*, 13 January 1984, p. 9; "The Scramble for the Next Superdrug," *Fortune*, 19 October 1981, pp. 94 ff.

TABLE 2

REPORTED CASUALTIES FROM NEW DRUGS, 1950–80
(Accidents Involving 100 Major Casualties or More)

Drug	Period/Location	Effects	Deaths	Other Casualties
Thalidomide	1960–61/worldwide	Birth defects		>10,000
Isoproterenol (inhalers)	1962–69/U.K. et al.	Asthma deaths in children	3,500	
Clioquinol	1956–70/Japan	SMON victims[a]	500	5,700
DES[b]	1970/U.S.	Vaginal cancer in daughters	150	400
Practolol	1969–75/U.K.	Misc. claims[c]		1,000
Chloramphencicol	c. 1950/U.S.	ADRs from misprescription	753[d]	
MER-29	1960–62/U.S.	Misc. claims[e]		400–500
Cutter Vaccine	1955/U.S.	Polio cases	11	204
Aminorex	1966–68/Europe	Cases of pulmonary hypertension	>30	>300
Orabilex	1958–64/U.S.	X-ray patients poisoned	25–100	
Stalinon	1954/France	Encephalitis victims	110	

[a] 11,000 total victims including minor casualties. SMON = subacute myelo-optical neuropathy.

[b] Diethylstilbestrol. Death figure given represents ultimate casualties projected from data in A. L. Herbst, ed., *Intrauterine Exposure to Diethylstilbestrol in the Human*, Proceedings of 1977 Symposium on DES.

[c] Including 100 severe claims.

[d] Sam Peltzman's estimate; Peltzman op. cit., pp. 52–54.

[e] Fewer than 100 major settlements.

SOURCES: Testimony of Barbara Moulton in *The FDA's Process;* M. N. Dukes, ed., *Side Effects of Drugs*, 8th and 9th eds. (New York: Elsevier, 1975, 1980); Edward C. Lambert, *Modern Medical Mistakes* (Bloomington: University of Indiana Press, 1978).

or disabling casualties have been reported worldwide since 1950. It should be noted that the overwhelming number of casualties is from three major disasters, any one of which could have caused as many as 10,000 casualties in the U.S. population. In fact, all three occurred during the 1960s in foreign countries where regulation was less strict than in the United States. Avoidance of these accidents may thus be taken as a useful indication of the value of present FDA regulation.

In comparison, the number of drug casualties reported in recent years appears relatively insignificant. Although critics of FDA reform have pointed to several drug accidents since 1981, none of these is above the 100-casualty cutoff of Table 2: the most significant include Oraflex, which claimed 11 lives in the United States and 62 more in Great Britain; Selacryn, which caused 25 deaths; Zomax, which caused 5 deaths; and E-Ferol, a vitamin product that was marketed without FDA approval, which killed 38 infants. In sum, these casualties amount to less than one-twentieth of those casualties associated with any of the three major drug disasters involving thalidomide, isoproterenol, and clioquinol.

Thalidomide is the only example of a major disaster involving a drug for which the FDA actually denied approval. In 1961 an epidemic of birth defects was observed in West Germany among babies born of mothers who had taken the tranquilizer thalidomide during pregnancy. News of the disaster broke out after the FDA had postponed approving an NDA for thalidomide. Worldwide, some 10,000 babies were born with severely defective limbs and other birth defects attributable to thalidomide. Had thalidomide been approved in the United States, an additional 10,000 to 19,000 casualties might have occurred.[51]

The FDA's success in averting disaster, however, appears to have been largely fortuitous, given the scientific knowledge of the time. Had FDA ordered further laboratory studies, it seems quite possible that the problem with thalidomide would not have been found. Subsequent animal studies failed to reveal thalidomide's effects in 13 out of 22 animal experiments, including those conducted with the most commonly used laboratory species.[52] It therefore seems altogether possible that the FDA would have eventually approved thalidomide had it not been for the tragic reports based on human experience in Europe. Ironically, even though the thalidomide disaster was widely interpreted as evidence of the need for longer premarket review, it actually illustrated the basic unreliability of premarket studies. Had other nations adopted the same cautious policies as the FDA in reviewing thalidomide, the United States itself might well have suffered disaster.

---

[51] The estimate of 10,000 is based on worldwide incidence rates; 19,000 is Peltzman's worst-case estimate based on West German incidence rates. See Peltzman, p. 55.

[52] Robert L. Brent, "Drug Testing in Animals and Teratogenic Effects: Thalidomide in the Pregnant Rat," *Journal of Pediatrics* 64, no. 5 (1964): 762–70.

The other two major drug disasters in Table 2 both involved unusual usage problems with old drugs that were already on the market in the United States. In the late 1960s asthma inhalers containing the widely used drug isoproterenol were found to be causing an epidemic of deaths among children in Great Britain, Australia, and elsewhere. The cause of the reactions appears to have been an unusually concentrated dosage form (other kinds of isoproterenol inhalers were and still are available in the U.S. and elsewhere, delivering the drug in a less concentrated dosage that has turned out to be safe). Between 1962 and 1970, it is estimated that some 3,500 children died from isoproterenol overdoses.

A similar disaster in the United States might have resulted in some 8,000 deaths.[53] That this did not happen seems largely fortuitous. Although the more dangerous form of isoproterenol was never submitted for FDA approval, it is doubtful whether its hazards could have been detected in premarket testing. Adverse reactions to isoproterenol appear to have been specifically connected with excessive use, and they were never observed in a clinical setting.[54] Wardell and Lasagna concluded, therefore, that isoproterenol was "precisely the type of adverse reaction that could not be reliably detected and intercepted in the premarketing phase" of drug testing.[55] However, casualties in foreign countries might have been reduced had more attention been paid to early warnings; deaths were linked to isoproterenol as early as 1964, three years before the first public warnings were issued.

During the 1960s thousands of Japanese users of the antidiarrhea drug clioquinol were left crippled, blinded, or otherwise disabled by a nerve disease known as subacute myelo-optical neuropathy (SMON). As with isoproterenol, diagnosis was complicated by the fact that clioquinol had been widely used in other countries for many years with no major ill effects. The peculiarity of the Japanese experience has never been adequately explained. It may have been the result of ethnic or environmental factors unique to Japan, or to the fact that the drug was more routinely prescribed in heavy chronic dosages by Japanese doctors. The Japanese government recognized about 11,000 SMON victims, 4,700 of whom had filed damage claims as of 1979.[56] Scaled to the U.S. population, a similar disaster involving clioquinol would have claimed twice as many victims.

---

[53] Based on British mortality data in Paul Stolley, "Asthma Mortality: Why the U.S. Was Spared an Epidemic of Deaths Due to Asthma," *American Review of Respiratory Disease* 105 (1972): 883.

[54] P. J. D. Heaf, "Deaths in Asthma: A Therapeutic Misadventure?" *British Medical Bulletin* 26 (1970): 245.

[55] Wardell and Lasagna, p. 99.

The FDA played no direct role in the clioquinol tragedy. Clioquinol had been on the market for many years and, thus, had avoided the rigors of the modern new drug approval process. The FDA, however, had outlawed over-the-counter sales of clioquinol in 1961, nine years before the drug was banned in Japan, and it had been used cautiously in the United States for many years due to suspicions of toxicity.[57] It therefore seems likely that an FDA-style new drug approval review of clioquinol could have prevented the SMON outbreak in Japan.

In sum, only one major drug disaster, that involving thalidomide, was actually averted by FDA intervention, and this may well have been by accident. One other major disaster, that involving clioquinol, was probably preventable by tighter premarket approval procedures. The third disaster, involving isoproterenol, probably was not preventable.

Among the lesser drug accidents shown in Table 2, there is only one case—namely, that of practolol—in which it has been argued that FDA regulation might have played a role in preventing U.S. casualties.[58] Practolol, however, was never actually submitted for FDA approval, and its side effects were discovered only after 250,000 patient-years of experience in the British market.[59] In other instances, it is dubious whether any amount of premarket testing could have averted disaster. The antiobiotic chloramphenicol, for example, was discovered to cause fatal reactions in one out of 25,000–50,000 patients, an order of magnitude more that the number of subjects normally enrolled in clinical testing. Similarly, the effects of DES did not manifest themselves until nearly a generation had passed (however, DES would have failed the post–1962 requirement for efficacy testing).

---

[56] William Chapman, "A Japanese Tragedy, Dirodohydroxyquinoline," *Washington Post* 18 March 1979; reprinted in *The FDA's Process*, pp. 1288ff.; other information on clioquinol taken from M. N. Dukes, ed., *Side Effects of Drugs*, 8th ed. (New York: Elsevier, 1975), p.794; and Gilman, et al., pp. 1064–65.

[57] Testimony of Barbara Moulton, in *The FDA's Process*, pp. 425, 1228. Clioquinol has since been withdrawn from the market in most developed countries but continues to be exported to the Third World despite medical misgivings about its use. See Gilman et al., pp. 1064–65.

[58] Testimony of Barbara Moulton, in *The FDA's Process*, p. 425.

[59] Testimony of Matthew Connolly, in *The FDA's Process*, p. 303.

Thus, there have been only one or two major drug accidents that could have been averted through stricter premarket testing, and one or two that could not have been prevented. Altogether, the worldwide 30-year total of new drug casualties shown in Table 2 comes to about 5,100 fatalities and 18,000 disabling injuries, of which approximately 1,000 deaths and 1,000 injuries occurred in the United States. Had the United States experienced the same casualty rate as other developed nations, another 1,000 deaths and 8,000 injuries might have occurred—the rough equivalent of one major thalidomide-type disaster in two or three decades. This is about the rate at which other developed countries have experienced major disasters.

At this rate, the benefits of FDA regulation relative to that in foreign countries could reasonably be put at some 5,000 casualties per decade, or 10,000 per decade for worst-case scenarios. In comparison, it has been argued above that the cost of FDA delay can be estimated at anywhere from 21,000 to 120,000 lives per decade. These figures would seem to support the conclusion that the costs of post–1962 regulation outweigh benefits by a wide margin, similar to Peltzman's results of a 4:1 cost-benefit ratio for the 1962 amendments.

Given the uncertainties in the data, these results must be interpreted with caution, although it seems clear that the costs of regulation are substantial when compared to benefits. However, one conclusion that can be drawn with certainty is that the FDA fails its own criterion for public health: *the FDA's new drug approval system is in no way proven "safe and effective."* It is therefore worth considering alternatives to the present system.

The Case for Consumer Choice

To this point, it has been assumed that new drug approval is the only means of preventing new drug accidents. The risk of drug accidents, however, can also be limited by means of usage warnings. Ideally, warnings could be designed to protect consumers from unnecessary risks while providing access to potentially valuable unproven drugs for those consumers who want them. Unfortunately, there has been relatively little scientific research on the design and impact of consumer drug information, possibly because the present regulatory system has largely preempted research in this area. Nevertheless, an examination of the drug accidents cited in Table 2 strongly suggest that usage warnings would have been as effective as regulatory approval in preventing major new drug disasters.

In this respect, it should be noted that the majority of adverse drug reactions are associated with relatively prolonged use and heavy dosages.[60] This was true in the case of isoproterenol and clioquinol, as well as in numerous other cases (MER-29, Aminorex, practolol, Oraflex). In addition, the risk of adverse reactions is concentrated in certain especially susceptible populations: the elderly, people using other drugs, children, and pregnant women. It seems likely, therefore, that the risk of adverse reactions could be greatly reduced if consumers—especially those in the most susceptible populations—were cautioned against prolonged and heavy use of new drugs, and if physicians and patients were alerted to be on the lookout for unknown reactions.

An important shortcoming of the present drug system is that it has no mechanism for informing consumers of the risks associated with new drugs. Once a product has been approved by the FDA as safe and effective, it is essentially indistinguishable from other drugs on the market. Furthermore, prescription drugs are commonly sold with no patient warnings or usage instructions from the manufacturer; present law requires only that pharmacists supply a typewritten label with the physician's instructions from the prescription (for example, "Take once a day").[61] Patients accordingly are dependent on the advice of their physicians, who are notably unreliable in providing drug information and warnings.[62] A solution to this problem would be to provide consumer warnings via written prescription drug "patient package inserts." Such inserts have been advocated by numerous consumer groups and have been considered by FDA, although proposals to require them have been shelved by the Reagan administration.[63]

Another way in which new drug risks could be reduced is through better postmarketing surveillance and testing. Many drug hazards go undetected for a prolonged period of time, as suggested in Table 2. For example, casualties from both isoproterenol and clioquinol occurred over the better part of a decade. Substantial mortality savings, therefore, might be achieved by earlier detection of adverse reactions. The United States has often been criticized for inadequate postmarketing surveillance, especially relative to other countries such as Great Britain and New Zealand.[64] It may be noted that the British surveillance system appears to have failed in the case of isoproterenol. This case, however, was a particularly difficult disaster, and there is little reason to believe that any other system, including premarket testing, would have worked any better.

---

[60] Inappropriate dosages have been found to account for as many as 80 percent of all drug accidents, including those attributable to familiar "old" drug reactions. See K. L. Melmon, L. B. Sheiner, and B. Rosenberg, "Medical Benefits and Risks Associated with Prescription Drugs: Facts and Fancy," *New England Journal of Medicine* 284 (1971): 1361, reprinted in Robert B. Helms, ed., *Drug Development and Marketing* (Washington, D.C.: American Enterprise Institute, 1975), pp. 5–13.

[61] Additional information is sometimes required by state law.

[62] According to a recent FDA survey, 80 percent of patients received no information from prescribing physicians. See the *New York Times*, 26 Feb. 1983, p. I–48. Physicians themselves are often woefully ignorant about the drugs they prescribe. See, for example, Milton Silverman and Philip R. Lee, *Pills, Profits, and Politics* (Berkeley: University of California Press, 1974), pp. 282–304; and Temin, *Taking Your Medicine*, pp. 88–119.

[63] It seems likely that modern product liability laws would have obliged prescription drug manufacturers to provide patient package inserts had FDA regulations not specifically relieved them of this obligation. See Robert Temple, "Legal Implications of the Package Insert," *Medical Clinics of North America* 58, no. 5 (September 1974): 1151–59.

For other major drug accidents, it seems likely that a combination of consumer warnings and postmarketing surveillance would have been roughly as effective as new drug approval in averting casualties. In the case of clioquinol, it seems likely that a British-style surveillance system in Japan would have detected the unusual symptoms of SMON poisoning early on and greatly reduced the number of casualties, although casualties might have been even lower if detected in premarket testing. Moreover, insofar as clioquinol's toxicity was suspected at the time, warnings against overdosage would have been appropriate.

In the case of thalidomide, the argument for warnings is even stronger, given the likelihood that premarket testing could not have detected the hazard of birth defects. Even today, animal testing is considered to be a remarkably unreliable guide to teratogenicity in humans. It would therefore seem hazardous to rely on premarket testing as a strategy for preventing future thalidomide-like disasters. There seems little question though that the risk of thalidomide-like casualties could be greatly reduced by warning pregnant women and their physicians to be cautious when using new drugs that have not specifically been tested for their effects during pregnancy. The presence of such a warning may actually be the reason why a major thalidomide-like disaster has not recurred. In fact, the FDA commonly approves drugs without any testing in pregnant women, but there is a specific warning to this effect on the package inserts, and physicians and women are widely aware of the need to be cautious in using new drugs during pregnancy.

It seems likely therefore that the risk of new drug accidents could be controlled as well by informational warnings as by strict premarket approval standards. While further behavioral research on consumer drug information is necessary, existing evidence strongly suggests that informational warnings could be designed to assure informed consent even for the most dangerous drugs.[65] Studies of patient package inserts already have revealed a strong consumer interest among people of all educational backgrounds in written prescription drug information.[66] It is also well known that drug warnings can be overdesigned so as to frighten patients away from beneficial medication.[67] In addition, given the publicity accorded to new drug accidents, there is reason to believe that the public is sensitive to their danger. In a recent poll, for example, only 18 percent of the respondents thought there is "very little" danger in taking new drugs, 38 percent said there is "great" danger, and 48 percent said there is "some danger."[68] Finally, a striking demonstration of the feasibility of informed choice in drugs is now taking place in Great Britain, where thalidomide has been administered to women of childbearing age for treatment of a rare condition known as Behcet's Syndrome.[69] This has been done by having the women sign a statement affirming that they have been warned of the drug's dangers and that they do not intend to get pregnant during treatment.

---

[64] Wardell and Lasagna, pp. 97–107.

Further research is necessary to determine the best design of a consumer drug warning system. However, an informed choice system would most likely feature the following four measures:

1. Written patient package insert warnings with prescription drugs.
2. Improved postmarketing surveillance of adverse drug reactions.
3. A system of graded safety and efficacy ratings for unproven drugs instead of the present simplistic categories of approved/not approved "safe and effective."

---

[65] Noninformational measures could arguably be necessary to prevent drug misuse (1) in the case of antibiotics, overuse of which may have deleterious third-party effects in breeding resistant organisms, and (2) in the special case of psychoactive or addictive drugs (for example, tranquilizers, painkillers, amphetamines, and illegal narcotics), which pose unique problems in an informed-choice framework because they appear to challenge the fundamental assumption of consumer free choice. The latter are now strictly regulated under the Controlled Substances Act, a full treatment of which is beyond the scope of this article. Nevertheless, it seems posible that the risk of accidental addiction could be satisfactorily minimized with no further legal restrictions other than perhaps some limitation on over-the-counter sales.

[66] Readership of patient package inserts has been reported at 70 percent and higher for first-time drug users, independent of educational level. See D. E. Kanouse, S. H. Berry, B. Hayes-Roth, W. H. Rogers, and J. D. Winkler, "Informing Patients About Drugs: Summary Report on Alternative Designs for Prescription Drug Leaflets," prepared for FDA (Santa Monica, Calif.: Rand Corp., August 1981), p. 3.

[67] Ibid., pp. 32–33; see also L. C. Epstein and L. Lasagna, "Obtaining Informed Consent," *Archives of Internal Medicine* 123 (June 1969): 682–88.

[68] Trendex survey for General Electric, cited in *Public Opinion* (June/July 1983): 31.

[69] *San Francisco Chronicle,* 3 October 1983, p. 3.

4. Special precautions in the case of especially risky or unproven drugs, such as prominent oral and/or written warnings explicitly acknowledged by patients and/or physicians with written signatures.

Given the preceding measures, it would seem possible to abolish mandatory drug approval entirely in favor of an informed choice policy. One way this could be done is to make unapproved drugs available with the explicit warning "NOT FDA APPROVED," and with whatever other warnings the FDA felt appropriate.[70] This would provide access to unapproved drugs for those who wanted them while maintaining the present level of FDA protection for those who did not. It would even be possible to raise FDA standards to a higher level than presently by establishing a new category of "especially safe and effective" drugs, while allowing other drugs to be sold with additional safety warnings. Moreover, it seems possible that drug warnings could be provided more efficiently through private-sector institutions such as medical advisory boards, health insurers, or consumer groups, than via government agencies, which may lack adequate incentives for providing consumers with the relevant information.[71]

An information-oriented drug policy offers certain advantages over the present approval system. The proposed drug policy would allow new drugs to be introduced more quickly, thereby reducing the costs of the present drug lag, and it would provide warnings so that only those consumers who preferred a new drug would be exposed to the increased risk. It has been argued above that the public health benefits of reduced regulatory delay would probably far outweigh costs, by as much as a few thousand lives per year. Although the proposed policy would probably entail some increase in the total number of new drug casualties, due to greater use of experimental and high-risk drugs, it is conceivable that an overall reduction in new drug casualties could result. This overall reduction could stem from the better information available to consumers about potential new drug risks. More important, drug label warnings would give consumers better information about the hazards of existing, "old" prescription drugs that are now on the market. Since present evidence suggests that the overwhelming number of drug accidents are due to old, not new, drugs,[72] it seems quite possible that there would actually be a substantial reduction in the total number of drug-induced injuries under an informed choice system. Furthermore, those casualties that did occur, in general, would be limited to people who had been specifically informed of the risks, which often is not the case under the present system. Finally, informed choice would encourage a more open attitude toward the use of experimental and unproven drugs, increasing medical knowledge for the benefit of everyone.

---

[70] For similar proposals, see Durk Pearson and Sandy Shaw, *Life Extension* (New York: Warner Books, 1982), pp. 590 ff.; and Murray Weiner, "Should the Public Have the Legal Right to Use Unproven Remedies? Yes," in Louis Lasagna, *Controversies in Therapeutics* (Philadelphia: W. B. Saunders, 1980).

[71] Political pressures may make it difficult for public agencies to act quickly and fairly, and action may be delayed through legal appeals. The FDA has taken some 20 years to review over-the-counter drug efficacy with comparatively little effect on improving consumer drug education. Similarly, FTC regulation of drug advertising has been notoriously ineffectual. See Bruce Yandle, "The Cost of Getting Nowhere at the FTC," *Regulation* (July/August 1981): 43–47.

Conclusion

There appears to be little justification for the present FDA approval system. From a public health standpoint, the FDA's post–1962 approval policies appear to be neither safe not effective. Reforms to speed up new drug approval may mitigate, but not eliminate the basic contradictions of a system based on societywide approval of safety and efficacy.

The evidence suggests that drug consumers can be protected as well by informational warnings as by restrictions on market choice. The most promising alternative appears to be an informed choice drug system, where individual consumers and physicians could choose for themselves what drugs to take given appropriate informational warnings.

---

[72] Wardell and Lasagna, pp. 100–3.

LABOR LAW AND THE FIRST AMENDMENT

Charles W. Baird

Lane Kirkland, the president of the AFL-CIO, recently claimed that the reason why labor unions are currently unpopular is "for the same reason journalists are unpopular. We both assert our First Amendment rights."[1] In a 1984 monograph[2] I argued that American labor unions enjoy more respectability than they deserve because most of the conventional wisdom regarding labor union history is in fact little more than myth. In this article I explain why I think Mr. Kirkland's suggestion that labor unions and the First Amendment go well together also claims more respectability for American unionism than the facts justify.

Two First Amendment freedoms are at issue: freedom of speech and freedom of association. The Labor Management Relations Act (LMRA), which is the original 1935 Wagner Act as amended by the 1947 Taft-Hartley Act, the 1926 Railway Labor Act as amended in 1951 (RLA), and state laws which affect government employees in their jurisdictions all have essentially the same anti-First Amendment provisions. The RLA governs unionism in railroads and airlines, the LMRA governs unionism in the rest of the private sector, and the state laws govern public sector unionism.

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Professor of Economics at the California State University at Hayward. An early version of a portion of this paper was presented at the monthly Library Seminar at the Institute for Humane Studies. The author thanks the seminar participants for their helpful comments. He also wishes to thank Gregory Christainsen for reading the penultimate draft and making several suggestions for improvement.

[1] *San Francisco Examiner*, 31 March 1985, p. A12.

[2] *Opportunity or Privilege: Labor Legislation in America*, The Social Philosophy and Policy Center, Bowling Green State University, Bowling Green, Ohio, 1984.

Freedom of Speech

This basic freedom has been put in jeopardy by American labor law and its administration in at least two ways: the use of compulsory agency fees paid by nonmembers to propagate partisan and ideological views with which those nonmembers disagree, and the imposition of explicit restrictions on the speech of employers during certification elections.

Use of Nonmember Agency Fees

In 29 states under the LMRA, and in all states under the RLA, unions are permitted to negotiate with employers for the inclusion of union security clauses in collective bargaining contracts. One form of union security is called the agency shop wherein a worker does not have to join a union in order to continue to work, but he or she must pay "agency fees" or "service fees" in order to do so. Typically these fees have been nearly the same as the dues that regular union members pay.

Now it should not come as a surprise to anyone, especially in light of last year's presidential election campaign, that labor unions spend a fair amount of their revenue from dues and agency fees for partisan and ideological purposes. While dues money collected from voluntary members may, without any impairment of their First Amendment freedoms, be so spent, compulsory agency fees paid by nonmembers may not be. For example, if the faculty union at California State University had an agency shop collective bargaining contract (which it does not) I would have to pay money to the National Education Association (NEA) in order to keep my job (in spite of the fact that I have academic tenure). The NEA used its revenues to speak out forcefully in voice and print in favor of Walter Mondale's candidacy for president, a political endorsement I did not share. If any of that revenue came from compulsory agency fees extracted from me, my freedom to choose the political candidate of whom I wish to speak out in favor would be impaired.

Thomas Jefferson, with reference to the question of the establishment of a tax-supported religion, wrote: "To compel a man to furnish contributions of money for the propagation of opinions which he disbelieves, is sinful and tyrannical."[3] If compulsory agency fees had existed then Jefferson almost certainly would have said the same thing in regard to them.

---

[3] "Virginia Act for Establishing Religious Freedom" (1784) in *The Writings of Thomas Jefferson* (Washington, D.C.: Thomas Jefferson Memorial Association, 1903), vol. 2, pp. 300–3.

The U.S. Supreme Court has for long recognized this threat to First Amendment freedoms, but until very recently it was practically impossible for nonmember agency fee payers to get relief from the use of their money by unions for partisan political and ideological advocacy.

The first Supreme Court case on this issue was *Machinists v. Street* [367 US 740 (1960)], wherein the Court unequivocally stated that the use of compulsory agency fees for partisan political or ideological purposes was an unconstitutional impairment of the freedom of speech. Revenue from compulsory fees could only be used for nonmembers' shares of the costs of collective bargaining, contract administration, and grievance adjustments. Unfortunately the Court did not specify any remedies available to aggrieved nonmembers. How much money should be refunded? How should that be determined? Would nonmembers have to pay amounts equal to regular dues and then apply for refunds, or should future collections from nonmembers be reduced to begin with? Since these and related questions were left unanswered, the Court's decision had little practical impact.

Then in *Railway Clerks V. Allen* [373 US 113 (1963)] the Court suggested a "practical decree" by which the proper reduction of agency fees below ordinary dues might be determined. This was only a suggestion, however, and the Court left it up to the parties involved to work out their own solutions. Again, very few people paid any attention to the problem.

The next Supreme Court case, *Abood v. Detroit Board of Education* [431 US 209 (1977)], involved a public sector union. One issue in the case was the constitutionality of an agency shop contract in public employment. The Court decided that, just as in the private sector, an agency shop in public employment is constitutionally permissible. (I will discuss this issue below.) The other issue in the case was the permissible uses of compulsory agency fees paid by nonmembers. Here the Court, consistent with both *Street* and *Allen*, declared that the teacher union could not use such fees for anything other than collective bargaining, contract administration, and grievance adjustment. The fact that government employment was involved did not change matters on this point. However, the teacher union had adopted an internal remedy whereby the aggrieved parties could seek restitution, so the Court deferred "further judicial proceedings pending the voluntary utilization by the parties of that internal remedy as a possible means of settling the dispute."[4]

---

[4] Case "Syllabus," p. III.

The most recent, and the most significant Supreme Court decision on this issue was announced on 25 April 1984. In *Ellis v. Railway Clerks* [No. 82-1150], the Court was very specific about the permissibility of certain uses of compulsory agency fees paid by nonmembers and about the propriety of rebate procedures available to aggrieved parties. It reiterated its judgment that the use of such fees for partisan political or ideological purposes of any sort was prohibited and that the use of such fees for collective bargaining, contract administration, and grievance adjustment was permitted. But the Court went on to consider six other uses of such fees: (1) a union's national convention, (2) litigation not involving collective bargaining or grievance settlement, (3) union publications, (4) social activities, (5) death benefits for employees, and (6) general organizing efforts.

The Court declared all expenditures in categories (2) and (6) to be impermissible. As examples of impermissible expenditures in category (2) the Court stated that:[5]

> objecting [airline] employees need not share the costs of the union's challenge to the legality of the airline industry mutual aid pact; of litigation seeking to protect the rights of airline employees generally during bankruptcy proceedings; or of defending suits alleging violation of the non-discrimination requirements of Title VII.

With regard to expenditures for general organizing efforts the justices said:[6]

> Using dues exacted from an objecting employee to recruit members among workers outside the bargaining unit can afford only the most attenuated benefits to collective bargaining on behalf of the dues payer.

Expenditures in category (1) unrelated to electing officers, establishing bargaining goals and priorities, and formulating overall union policy, as well as expenditures in category (3) related to political causes were also declared to be impermissible uses of agency fees. The only category in which all expenditures were declared to be permissible uses of such fees was category (4). The Court declared the issue moot with regard to category (5) since the respondent union was no longer the bargaining agent for the plaintiffs.

The Court went further. It declared that unions collecting compulsory agency fees had to set those fees below regular dues. A union can no longer collect amounts equal to regular dues from nonmembers and then give rebates, even if it pays interest on the amount to be rebated. The Court stated that such a rebate scheme amounts to "an involuntary loan for purposes to which the employee objects."[7]

---

[5] Slip opinion no. 82-1150, at 17.

[6] Ibid., at 15–16.

The final case of interest on this issue has not yet been heard by the Supreme Court. In *Beck v. Communications Workers of America* [112 LRRM 3069 (1983)] a U.S. District Court judge in Baltimore ruled that the union had to refund 79 percent of the fees collected from nonmembers between 1 January 1976 and 31 March 1983 because it had used that portion of the fees for impermissible purposes. The case is currently on appeal and may or may not eventually be heard by the Supreme Court. In light of the *Ellis* decision the union may decide to concede. However, since *Ellis* involved the RLA and *Beck* involves the LMRA that remains to be seen.

If Lane Kirkland's suggestion that the AFL-CIO approves of, supports, and acts upon the basis of First Amendment freedoms were true I would expect him to applaud the *Ellis* and *Beck* decisions. To no one's surprise he has not.

Restrictions on Employers' Speech in Certification Elections

It is commonly believed that prior to the passage of the original Wagner Act in 1935 most American workers were the helpless victims of their employers who tried every trick they could think of to prevent workers from forming unions. Thus one important aspect of the LMRA is its provisions which are designed to guarantee that employers will in no way prevent workers from voting to be represented by unions.

Section 7 of the LMRA states:

> Employees shall have the right to self-organization, to form, join, or assist labor organizations, to bargain collectively through representatives of their own choosing, and to engage in other concerted activities for the purpose of collective bargaining or other mutual aid or protection. . . .

Section 8(a)(1) states:

> It shall be an unfair labor practice for an employer to interfere with, restrain, or coerce employees in the exercise of the rights guaranteed in Section 7.

The courts and the National Labor Relations Board (NLRB) have interpreted this to mean that certification elections must be held under "laboratory-like" conditions so that the genuine free choice of workers will be revealed. Workers are assumed to have very fragile wills which can easily be thwarted even by the slightest hint of threat of reprisal or promise of benefit from employers trying to discourage a pro-union vote. In spite of a definitive study published in 1976 that demonstrated, both logically and empirically, that workers are capable of weighing the merits of competing campaign claims by employers and unions,[8] the NLRB and the courts continue to treat workers as children who must, even at the expense of First Amendment freedoms, be protected from "unfair" campaign speech from employers. Workers are assumed to be able to weigh the competing campaign claims of candidates for the Congress or for the presidency, but not those of employers and unions.

---

[7] Ibid., at 7.

There are two Supreme Court decisions that guide the NLRB and lower courts in this matter. In *NLRB v. Exchange Parts Co.* [375 US 405 (1964)] the employer, during a certification election campaign, announced certain changes of policy that were beneficial to workers. Specifically the employer announced (1) the establishment of a new floating holiday which the workers could elect to take on their birthdays, (2) a new system for computing overtime during holiday weeks, and (3) a new vacation schedule.

The Court ruled that these announcements were violations of Section 8(a)(1) of the LMRA and set aside the anti-union election results. The Court explicitly recognized that First Amendment speech rights were involved, but asserted that the government's interest in peaceful labor relations were overriding. In the words of the Court (at 409):

> The danger inherent in well-timed increases in benefits is the suggestion of a fist inside the velvet glove. Employees are not likely to miss the inference that the source of benefits now conferred is also the source from which future benefits must flow and which may dry up if it is not obliged.

The second Supreme Court case was *NLRB v. Gissel Packing Co., Inc.* [395 US 575 (1969)]. This case involved a firm that had previously suffered a long strike which had closed down two plants for three months in 1952. The plants were subsequently reopened on a non-union basis. Then in 1965 the Teamsters began an organizing campaign among the employees. During that campaign the employer stated that employees who signed Teamster authorization cards were forgetting the lessons of the 1952 strike that had caused employees as well as the employer a lot of grief. The employer also stated that the firm was not financially strong and therefore any strike could result in a plant closing. Moreover, the employer went on to say, many employees, because of age and lack of education, would have a difficult time finding alternative employment.

---

[8] Julius G. Getman, Stephen B. Goldberg, and Jeanne B. Herman, *Union Representation Elections: Law and Reality* (New York: Russell Sage Foundation, 1976).

The Court found that these statements by the employer also violated Section 8(a)(1) of the LMRA. The statements were merely "threats of reprisal" which are not protected by the First Amendment. Such threats of reprisal make it impossible for workers to exercise genuine free choice in making up their minds how to vote. Thus the Court ordered that the employer recognize the union as the exclusive bargaining agent of the employees even though the employees voted against such representation. This "bargaining order" was presumably imposed on the grounds that any subsequent election would be tainted by the threats of reprisal, and so would be invalid.

Since this seems to be such an obvious infringement of the employer's First Amendment rights and since the bargaining order remedy seems so coercive, it is worthwhile to consider the reasoning of the Court. In its words (at 617–18):

> Any assessment of the precise scope of employer expression, of course, must be made in the context of its labor relations setting. Thus an employer's rights cannot outweigh the equal rights of the employees to associate freely. . . . And any balancing of those rights must take into account the economic dependence of the employees on their employers, and the necessary tendency of the former, because of that relationship, to pick up intended implications of the latter that might be more readily dismissed by a more disinterested ear. . . .
>
> Thus, an employer is free to communicate to his employees any of his general views about unionism or any of his specific views about a particular union, so long as the communications do not contain a 'threat of reprisal or force or promise of benefit.' He may even make a prediction as to the precise effects he believes unionization will have on his company. In such a case, however, the prediction must be carefully phrased on the basis of objective fact to convey an employer's belief as to demonstrably probable consequences beyond his control. . . . If there is any implication that an employer may or may not take action solely on his own initiative for reasons unrelated to economic necessities and known only to him, the statement is no longer a reasonable prediction based on available facts but a threat of retaliation based on misrepresentation and coercion, and as such without the protection of the First Amendment.

In my judgment the employees' equal right to associate freely was abridged by the bargaining order remedy imposed by the Court. The Court stripped the employees of their voting rights under the LMRA and forced them to associate with the union on the flimsy grounds that employees' "economic dependence" on the employer made it impossible for them to come to a reasoned judgment on their own.

This economic dependence argument betrays the Court's misunderstanding of the true nature of the relationship between employees and employers in the labor market—namely, a contractual relationship based on voluntary exchange. Employees and employers are economically dependent on each other. Moreover, to secure the services of an employee, an employer must make an employment offer that is at least as good as any alternative offer the employee may receive. To obtain employment with a particular employer, an employee must make an offer that is at least as good as those made by other potential employees seeking the same employment. Both parties are dependent on their alternatives for their bargaining leverage. The common belief that employers' alternatives are always better than those of employees is not borne out by the facts. Monopsony is extremely rare.

The distinction between carefully phrased predictions, based on objective facts beyond the employer's control, about the effects unionization will have on a company on the one hand, and suggestions that unionization may make the employer, for reasons known only to him and which are unrelated to "economic necessities" decide to close a plant on the other, is one of mind-slipping wetness. Is not the employer the sole legitimate judge of what economic circumstances are sufficient to cause him to close? Who is to decide what is or is not an "economic necessity"? The *Gissel* decision in practice makes it very hazardous for an employer to even suggest that unionization of a plant will result in its partial or complete closing, for a union can always argue that any reasons offered by the employer for such action are insufficient to constitute "economic necessity."

The *Exchange Parts* and *Gissel* decisions together grant unions a competitive advantage in organization campaigns and certification elections. It is these decisions that permit unions to argue that an employer (for example, J. P. Stevens & Co.) who speaks out against unionization, or who attempts to forestall unionization by offering his employees better terms and conditions of employment during an organizing campaign and/or a certification election, is guilty of "repeated violations of labor law." All such employers are really guilty of is exercising their First Amendment rights. Not content with the Court-granted competitive advantage they already have, labor unions tried, but failed, to get even more advantages over employers and employees who want to stay away from unions by the so-called Labor Law Reform Bill of 1978.

Freedom of Association

Although the Supreme Court thinks that American labor law is aimed at assuring workers of their freedom of association and is, therefore, to be put in the balance against employers' freedom of speech, I contend that some of the main provisions of American labor law, especially those of the LMRA, effectively deny freedom of association to many workers. I have four specific provisions in mind: exclusive representation, union security, the proscription of company unions, and mandatory bargaining.

Exclusive Representation

Section 9(a) of the LMRA provides that a union that wins a representation (certification) election is the exclusive bargaining agent for all of the workers in the bargaining unit. A union that wants to organize the workers at a plant begins by trying to secure the signatures of at least 30 percent of those workers on authorization cards. A person who signs such a card declares that he or she wants the representation services of the union. Once at least 30 percent of the workers have signed, the union then requests the NLRB to set up and administer a secret ballot representation election. If only one union is attempting to become the exclusive bargaining agent the ballot will be a choice between the union and "no representation" (that is, no union). The winner is that choice that gets 50 percent plus one or more of the votes cast. If more than one union is seeking to be the exclusive bargaining agent the ballot will be a choice between each of the unions in contention as well as no representation. The winner is again that choice (if any) that receives 50 percent plus one or more of the votes cast. Should there be no winner, there is a runoff between the two choices that received the most votes cast.

The union that becomes the exclusive bargaining agent gets to represent all the workers in the bargaining unit. It represents those who wish its representation services, and it represents those who do not want its services. The latter group consists of workers who want the representation services of other unions and those who do not want the representation services of any union. I claim that the freedom of association of this group is denied. Individuals therein are forced to associate with the winning union against their will. Surely the freedom of association includes the right to choose to abstain from associating with any particular person or group. Forced association is not free association.

Those who support the principle of exclusive representation do so by analogy with congressional elections. An exclusive representative to the House of Representatives is chosen by secret ballot elections in every congressional district. A winner has only to get a plurality of the votes cast in order to represent everyone in his or her district, including those who voted for losing candidates as well as those who did not vote at all. The U.S. Constitution specifically provides that that is how things ought to be, and no one demurs. Since such a procedure is acceptable in the case of congressional elections, the argument goes, it must also be acceptable in the case of union representation elections. Anyone who does not approve of such a procedure must be anti-democratic.

But that argument is invalid. It is invalid because the government-union analogy is invalid. Unions are not governments. A government is a natural monopoly. Indeed, the definition of government is that it is the agency within a political jurisdiction that has a monopoly on the legal use of force. There cannot be competition between two or more governments in a political jurisdiction. The name for such competition is war. A union, on the other hand, is a private association of individuals. It is a club like the Rotary or the Lions. There is nothing natural about monopoly in the case of private clubs. Individuals are usually free to choose whether to associate with a private club at all, and that association cannot be compelled by any majority vote. Moreover, it is natural for private clubs to coexist and compete with each other for allegiance and members. An American who exercises the freedom to choose which, if any, private club with which to offer to associate, notwithstanding the outcome of any voting among others, is not being anti-democratic. He or she is merely exercising the freedom of association that our democratic *government* is elected to protect. The fact that the government took away that freedom when it adopted exclusive representation in the Wagner Act merely says that the democratically elected government acted unconstitutionally.

Of course, since *Marbury v. Madison* (1803) it is the Supreme Court that gets to decide what is and what is not constitutional, and in *NLRB v. Jones and Laughlin Steel Corporation* [301 US 1 (1937)], by a vote of five to four, the Court declared the Wagner Act was constitutional. But that decision, long referred to as the "switch in time that saved nine," says more for the political astuteness of the Court than it does for its constitutional reasoning. Just one year earlier, in *Carter v. Carter* [298 US 238 (1936)] the Court unanimously threw out the Guffey Coal Act citing the exclusive representation features of that law as one of the reasons for its decision. But on 5 February 1937, four days before the Court heard oral arguments on the Wagner Act, President Roosevelt announced his infamous court-packing plan in which he would have increased the size of the Court by appointing new members who would be more amenable to such New Deal measures as the Wagner Act. Five justices were thereby induced into changing their views regarding the constitutionality of compulsory exclusive representation applied to private associations.

Union Security

Section 8(a)(3) of the LMRA empowers labor unions that have exclusive bargaining agent status to negotiate with employers for the inclusion of union security provisions in collective bargaining contracts. Three types of union security arrangements are permitted: union shop, agency shop, and maintenance of membership. In a union shop the employer can hire anyone he or she chooses; but, after a 30-day probationary period, a new employee must join the union as a condition of continuing employment. In an agency shop employees do not have to join the union, but they have to pay service (agency) fees to the union as a condition of continuing employment. Under a maintenance of membership agreement an employee who is a member of the union at the inception of a collective bargaining contract must, as a condition of continuing employment, continue to be a member until that contract expires.

Compelling a person to pay dues or service fees to a private club as a condition of continuing to work is an egregious denial of the freedom of association. Association through the wallet is perhaps the most significant association of all. If a person is not free to choose to refrain from paying money to a private organization he is forced to associate with that organization. That organization gets a prior claim to income earned by the dissenting individual—that is, it gets the power to tax.

Supporters of union security arrangements justify them by the free rider argument. Since the LMRA requires an exclusive bargaining agent to represent all workers in the bargaining unit, it is only fair that all such workers pay their fair share of the costs of such services. Without a union security arrangement it would be possible for a worker to be a free rider—that is, to receive the benefits of union representation without paying anything for them. Simple justice requires that each person who receives the services must pay for them. A free rider would receive benefits at the expense of other workers who would have to pick up the free rider's fair share.

The obvious counterargument is that two wrongs do not make a right. While it is true that an exclusive bargaining agent must represent all workers, that provision of the law was not forced on unwilling unions. The unions fought a determined battle to get exclusive representation written into the law.[9] The solution to the free rider problem is to repeal exclusive representation. If unions represented only workers who freely chose on an individual basis to associate with them and support them, there could be no free riders.

---

[9] *Opportunity or Privilege,* chaps. 2 and 3.

If Lane Kirkland were really interested in supporting First Amendment freedoms for everyone he would work toward the repeal of exclusive representation. But because he is so wedded to the principle of exclusive representation he must fall back on union security to cope with free riders. Thus he uses one infringement of the freedom of association (exclusive representation) to justify a further infringement of that freedom (union security).

Not surprisingly, the U.S. Supreme Court does not agree that union security arrangements are egregious infringements of First Amendment freedoms. In *Railway Employees' Department v. Hanson* [351 US 225 (1956)] the Court disagreed with the Nebraska Supreme Court which argued that a union shop clause affecting Nebraskan railroad employees infringed the First Amendment of the U.S. Constitution as well as Nebraska's proscription of such arrangements within its borders. The U.S. Supreme Court recognized a valid First Amendment question, but declared (at 233):

> Industrial peace . . . is a legitimate [legislative] objective; and Congress has great latitude in choosing the methods by which it is to be obtained. The choice by the Congress of the union shop as a stabilizing force seems to us to be an allowable one.

The Court says the benefits of increased industrial peace are more than enough to compensate for the cost of the infringement of First Amendment freedoms. That proposition is debatable on the face of it, but when it is realized that the adoption of the original Wagner Act was the occasion for decreased industrial peace[10] the Court is seen as justifying the imposition of a cost with which there is no associated benefit.

Proscription of Company Unions

Section 8(a)(2) of the LMRA states:

> It shall be an unfair labor practice for an employer to dominate or interfere with the formation or administration of any labor organization or contribute financial or other support to it. . . .

This provision of the law is aimed at employer-sponsored unions which, according to conventional wisdom, were used by employers, prior to the adoption of the Wagner Act in 1935, to meet the requirements of Section 7(a) of the National Industrial Recovery Act (NIRA) while still denying their employees legitimate unionization rights. Under the NIRA workers were free to choose which sort of union, if any, with which to associate—an "independent" union such as the AFL, or an employer-sponsored union. The independent unions alleged that employer-sponsored unions (company unions) were simply fronts used by employers to mask their exploitation of workers. Workers who chose to associate with company unions were excoriated by the independent unions as cowards and lackeys.

---

[10] Ibid, chap. 4.

Between 1933 and 1935 independent unions launched a large number of recognition strikes. The purpose of these strikes was to force employees to disassociate from company unions and join "real" unions. According to folklore these strikes were battles of workers against employers, but in fact they were battles between independent unions and workers who chose to join company unions or chose to abstain from unionism altogether. The typical tactic of an independent union was to get a few employees at a plant to join up and then call a strike. The great majority of workers who wanted to continue to work were then harassed by "flying squadrons" of nonemployee pickets who were brought in from outside the local community to shut the plant down.

In any event, the Wagner Act settled the matter by proscribing company unions. A measure of "labor peace" was secured simply by denying one side of a dispute its rights to defend itself. This is the "peace" that comes from oppression.

It seems self-evident to me that freedom of association implies that if some workers find company unions to be superior to independent unions as devices for pursuing their goals, they ought to be free to choose to join company unions. And in fact, many employer-sponsored unions prior to 1935 were superior to independent unions in that way.

Those employer-sponsored unions were early forms of what we today call "quality circle" arrangements wherein employers and employees cooperated in the search for and implementation of improvements in production processes which enabled the firms involved to be more effective competitors. Company-sponsored unions were structured on the natural complementarity between workers and management. Cooperation was seen as the best way to serve the interests of both groups. Independent unions, on the other hand, have always been predicated upon an adversarial relationship between workers and management. They have promulgated the view that warfare (class warfare?) best describes the natural state of affairs between labor and capital.

Today American smokestack industries are encouraged to emulate the Japanese style of labor-management relations in order to rescue themselves from extinction. If it had not been for the proscription of company unions in 1935, those early forms of quality circle arrangements could have evolved into effective forms of labor-management relations which even the Japanese would envy.

More important, Section 8(a)(2) is today standing in the way of needed changes in unionized American industries. In *NLRB v. Cabot Carbon Co.* [360 US 203 (1959)] the Supreme Court held that employer-created employee committees set up to confer with management on a regular basis on matters of mutual concern such as quality control, production innovation, and complaint settlement violated Section 8(a)(2) of the LMRA. This ruling stands to this day, and it has been interpreted very narrowly by the NLRB in its rulings. In a recent journal article the NLRB's position was stated in this way:[11]

> [T]he Board has been very consistent in its rather strict interpretation that Section 8(a)(2) prohibits all employer support which *could have* impinged upon employee free choice. There is no requirement that an *actual* interference be proven. More importantly, evidence that there was no interference but merely cooperation with employee free choice has seldom been successful to negate the Board's presumption that employer support is unlawful in all circumstances.

The most direct way to remove Section 8(a)(2) as a barrier to innovation is to eliminate it from the law. That would permit employees and employers to cooperate in experimenting with alternative production arrangements, and it would also permit company-sponsored employee organizations to compete directly with labor unions for membership and allegiance. The latter form of competition would encourage beneficial innovation in labor-management relations,[12] and it would restore some freedom of association to workers.

Mandatory Bargaining

Section 8(a)(5) of the LMRA states:

> It shall be an unfair labor practice for an employer to refuse to bargain collectively with the representatives of his employees. . . .

---

[11] Curtiss K. Behrens and James Sollenberger, "The National Labor Relations Act: A Potential Legal Constraint Upon Quality Circles and Other Employer-Sponsored Employee Committees," *Labor Law Journal*, December 1983, pp. 777–78 (emphasis in original).

[12] Charles W. Baird, "Labor Law and Entrepreneurial Discovery," *Austrian Economics Newsletter*, Summer 1985.

Section 8(d) states:

> For purposes of this section to bargain collectively is the performance of the mutual obligation of the employer and the representative of the employees to meet at reasonable times and confer in good faith with respect to wages, hours, and other terms and conditions of employment. . . .

One aspect of the freedom of association is the right to engage in voluntary exchange contracting with others. But if a contract is based on voluntary exchange all parties thereto must consent to enter into bargaining to begin with. If I am told that I must bargain with some person or group, I am forced into the bargaining association. That is another self-evident infringement of First Amendment rights.

Whether an employer would choose to bargain with a labor union if he or she did not have to would depend on the economic circumstances of the moment. If good substitutes for the workers who are represented by the union that wants to bargain were readily available on terms the employer found attractive, the employer would probably not choose to bargain with the union. That is the way competition works—the party that offers the best terms wins. The employer is not responsible for the existing conditions in the labor market; he or she merely reacts to them as they are. If the law prevents the employer from accepting the terms offered by nonunion workers by insisting that the employer bargains with the union, the law exploits some workers at the behest of other workers. That violates the principle that the law is supposed to treat everyone equally. The terms offered by nonunion workers have just as much moral merit as the terms offered by union workers, even if the former are called "scabs."

The duty to bargain has even been used to prevent a company from contracting out maintenance work after its existing collective bargaining contract with its unionized maintenance workers expired. In *Fibreboard Paper Products Corporation v. NLRB* [379 US 203 (1964)], the Supreme Court declared that contracting out was a mandatory subject of collective bargaining. This means that a firm cannot switch from in-house operations to outside contractors, even when the contract covering the in-house workers expires, without securing the consent of the union that represents those in-house workers. A contract for (say) two years gives union workers a legally enforced right to force employers to bargain regarding all future years.

Moreover, the duty to bargain "in good faith" has come in practice to mean that employers must make concessions to unions during collective bargaining sessions. An employer who steadfastly refuses to improve his offer is in jeopardy of being charged with an unfair labor practice. If a strike is subsequently called because of failure to reach agreement the strike could be designated as an unfair labor practice strike rather than an economic strike. After an unfair labor practice strike is settled an employer is legally forced immediately to reinstate all strikers. Any replacement workers that were hired during the strike must be dismissed to make room for the returning strikers. (In the case of an economic strike, unless the settlement specifies otherwise, strikers merely have first claim on the jobs if and when replacement workers leave them.) The duty to bargain in good faith thus becomes in yet another way an impediment to the freedom of nonunion (replacement) workers to associate through contract with employers.

Conclusion

Folklore has it that the Wagner Act was a long overdue measure designed to guarantee to American workers their constitutional right of freedom to associate in labor unions. But that right had never been denied. Workers have always been free to form labor unions and collectively to pursue better terms and conditions of employment than they could get individually. The only restriction placed on such voluntary unions was they could not use force to compel employers to accept their demands or to prevent nonunion workers from working. In many ways the Wagner Act is more accurately viewed as an infringement of the freedom of association which grants private associations of some workers the legal use of force over other workers and employers.

In August 1984 Lane Kirkland, for entirely different reasons, said that it is perhaps time to repeal all but the "most basic" of American labor laws. I am not sure what he means by "most basic," but in my judgment the time has come, on First Amendment grounds, to give serious consideration to the repeal of the LMRA, the RLA, and all similar state legislation.

DEREGULATING URBAN TRANSPORTATION

Robert Cervero

The nation's air, rail, trucking, and intercity bus industries have all been deregulated in recent years with generally bipartisan support. The lifting of controls over fares, routing, and market entry has taken place in the name of economic efficiency—getting government out of the friendly skies and busy highways of this nation so that private market forces can prevail. Strangely, however, deregulation has been largely limited to intercity and transcontinental modes of transportation. There is no reason, however, why the same principles cannot be succesfully applied to urban transportation as well.

Current Problems in Urban Transportation

Present-day urban transportation problems are all too familiar to most Americans. Many of our nation's cities are plagued with overcongested and undermaintained highways and financially crippled transit systems. Maintenance and restoration of major arteries and freeways have routinely been deferred over the past several decades, even though private automobile and truck travel has steadily increased since the 1950s. Few municipal bus systems in the United States today recoup more than 40 percent of their costs from fares, relying instead largely on government bailouts to make up the difference. Yet transit wage rates have outpaced the rate of inflation over the past 15 years (one of the few employment sectors to accomplish this feat), even though service quality and overall productivity have

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Assistant Professor of City and Regional Planning at the University of California at Berkeley. A condensed version of this paper appeared in the May/June 1984 issue of *Regulation*.

generally fallen during this period.[1] In many cities buses often operate at crush loads during rush hours. During off-peak hours, however, restrictive work rules prevent the hiring of part-time help and may cause the same buses to run nearly empty.

The cost of operating the nation's transit systems and reconstructing and maintaining urban roadways today stands at over $20 billion per year, comprising almost 5 percent of total local, state, and federal budgets.[2] (This ignores capital expenditures, which, by comparison, have generally stabilized in recent years). The financial burden of urban transportation is quickly overwhelming government treasuries at a time when public pressure is mounting to curb spending.

Federal largesse has been particularly generous to transit interests over the past decade. Federal subsidies to urban transit grew from $132 million in 1970 (consisting largely of small capital grants) to over $3.2 billion in 1982, an increase of 2,400 percent.[3] For the same period, federal expenditures on highways grew only 93 percent, from $4.38 billion to $8.45 billion, and much of the increase was eaten up by inflation.[4] This shift in national policy reflects a common belief among many people—be they environmentalists, political conservatives, or even pro-highway supporters—that a strong national urban transit system is needed. Public transit enjoyed broad ideological appeal throughout the 1970s, and was openly embraced by Republican and Democratic administrations alike as part of their national urban policy packages.

Despite the massive infusion of government aid to transit, there has been only a relatively small payoff. National ridership has remained fairly stagnant since 1970, amounting to about 6 billion passengers annually.[5] Several studies document that a large share of government subsidies has been consumed by higher labor costs, lower productivity, and the unbusinesslike expansion of services into low-density, suburban markets.[6] The availability of subsidy dollars, many argue, has lulled transit operators into a false sense of security, which in turn has resulted in lax management practices and overly generous wage settlements.[7] Most important, the underwriting of public transit services has discouraged the emergence and expansion of many alternative and generally higher quality urban transit services, such as taxis and private commuter buses. Indeed, transit's monopoly along many dense corridors of U.S. cities has helped institutionalize inefficient operating practices while squelching competition and innovation.

---

[1] See American Public Transit Association (APTA), *Transit Fact Book* (Washington, D.C.: American Public Transit Association, 1984), preliminary estimates; Don H. Pickrell, "Sources of Rising Operating Deficits in Urban Bus Transit," *Transportation Research Record* 915 (1983): 18–24; and John R. Meyer and Jose A. Gomez-Ibanez, *Improving Urban Mass Transit Productivity* (Washington, D.C.: Urban Mass Transportation Administration, U.S. Department of Transportation, 1977), pp. 12–15.

[2] See APTA; *Highway Statistics, 1982* (Washington, D.C.: Federal Highway Administration, 1982).

[3] APTA.

[4] See *Highway Statistics, 1982; 1980 Summary of UMTA's Transit Assistance Program* (Washington, D.C.: U.S. Department of Transportation, 1980); and *Highway Trust Fund 24th Annual Report* (Washington, D.C.: U.S. Department of Transportation, 1980).

[5] APTA.

Perhaps the major public policy lesson in urban transportation during the 1970s and early 1980s has been that money itself will not lure Americans out of their cars and into public transit or other forms of ridesharing. Despite low transit fares, expansive transit routing, carpool matching programs and publicly sponsored vanpools, Americans have resisted attempts to change their commuting habits. As Alan Altshuler notes, Americans have continued to pursue low-density, auto-reliant lifestyles despite public efforts to convince them they should do otherwise.[8]

It is argued in this paper that more could be done to improve the quality of urban transportation and perhaps abate the current fiscal shortfall through the process of deregulation than through almost any other policy strategy. Current problems have created an unprecedented opportunity for introducing various regulatory reforms. Decontrol would allow jitneys, shared-ride taxis, vanpools, and private commuter buses to compete for parts of the transit market now largely monopolized by bus and subway operators. A competitive marketplace would force operators to restructure their services and price them more rationally. The riding public would benefit not only from a wider array of available travel modes but also from a more integrated system of urban transportation. The deregulation of minimum parking requirements found in most local zoning ordinances could also help relieve inner-city congestion and promote the use of alternative travel modes. Likewise, revising subdivision regulations to allow parking limits on new developments and density levels supportive of public-transit investments could have substantial longer-term payoffs.

---

[6] The two most complete studies on the "leakage" of transit subsidies are Douglass B. Lee, Jr., "Evaluation of Federal Operating Subsidies to Transit," working paper (Cambridge, Mass.: Transportation Systems Center, 1983); and Don H. Pickrell, *The Causes of Rising Transit Operating Deficits* (Washington, D.C.: Urban Mass Transportation Administration, U.S. Department of Transportation, 1983).

[7] See John Pucher, "Effects of Subsidies on Transit Costs," *Transportation Quarterly* 37 (October 1982): 549–62; and Robert Cervero, "Examining the Performance Impacts of Transit Operating Subsidies," *Journal of Transportation Engineering* 110 (September 1984): 467–80.

[8] Alan Altshuler, *The Urban Transportation System: Politics and Policy Innovation* (Cambridge, Mass.: MIT Press, 1979), p. 377.

Perhaps it is deregulation's potential for spawning a rich mix of different types of urban transportation services that is most promising. Many Americans have only one or two alternatives when traveling other than by car—generally bus or taxi. Yet the past decade has taught us that the provision of fixed-route, uniform-quality services will not lure significant numbers of people out of their cars. There is tremendous diversity in travel preferences—some want fast, comfortable services and are willing to pay a premium fare for them, while others are satisfied to travel under slower, more congested conditions if they are given a break at the farebox. A wide array of service options clearly offers the greatest hope of enticing commuters to switch their mode of travel.

The importance of service features is well known within the transportation field. Studies consistently show that commuters are far more sensitive to the quality of transportation services than price levels—that is, they are most likely to change their travel behavior, and perhaps switch modes, given dramatic changes in travel times or comfort levels. In fact, several studies have found transit riders to be at least twice as responsive to service changes (such as more frequent scheduling of buses) than to lower fares.[9] Factors such as reliability of schedules, assurances of a seat, and availability of temperature control have proven to be key determinants of what modes travelers choose. Time spent walking to a bus stop, waiting, and transferring is particularly abhorred by commuters. It should come as no surprise, then, that private commuter buses and taxis provide these desired premium features to their customers and are making a profit while doing so.

What we need, then, is a wider assortment of urban transportation services targeted to meet the diverse needs and travel preferences of urban Americans. Regulatory reform could inject a much-needed dose of competition into the urban transportation sector, just as it has done with the nation's airlines, railroads, and highways, and thereby give rise to a wealth of travel alternatives to the automobile.

Legacy of Urban Transportation Regulations

The regulation of urban transportation in the United States has long embraced fundamental principles of public welfare economics.[10] Taxis, jitneys, and private buses have fallen under government control in the belief that their ascendancy would undermine common-carrier public transit services and threaten public safety.[11] Similarly, parking is often regulated for the purpose of averting potential traffic congestion. Even though these regulations are said to protect the broader "public interest," interference in the private marketplace all too often has created inefficiencies and market distortions where none previously existed.

---

[9] Patrick Mayworm et al., *Patronage Impacts of Changes in Transit Fares and Services* (Washington, D.C.: U.S. Department of Transportation, 1980).

The underlying economic rationale for regulating urban transportation modes has historically been that they are natural monopolies—that is, they enjoy economics of scale (where average costs decline as use increases), as do water and electric utilities. Under such conditions, a sole operator can most efficiently and inexpensively provide services. If numerous competitors were allowed into the urban transportation market, the argument goes, they would "skim the cream" by taking the most lucrative routes and leaving the unprofitable ones. However, a single transportation company would operate in the public interest by operating both money-making and money-losing services, a practice commonly called cross-subsidization. To ensure high levels of transportation services throughout an urban region, then, regulators argue that the public has an obligation to protect carriers from excessive competition and ensure that they receive an adequate rate of return.

A variety of controls on urban transportation have emerged under this regulatory banner. Entry restrictions have been placed on taxis in most U.S. cities, while jitneys, shared-ride taxis, and other private ridesharing ventures have generally been regulated out of existence. Both fare levels and rate structures also have come under government scrutiny. In addition, regulations have sought to ensure safety through controls on vehicle inspection, driver qualifications, and general operating practices.

The first regulation of urban transportation in the United States involved the granting of franchises to electric utility companies to operate streetcar and trolley lines around the turn of the century. Franchises often stipulated a nickel fare, which traction companies were more than willing to accept, given the opportunity to bring transit under the monopoly of the electric utility industry.[12] Even at this time, most streetcar lines were being subsidized, as transit was often part of an overall real estate development package. The principle of cross-subsidization was being fully exploited—electric utilities would willfully sustain losses operating streetcar services as long as they alone could extend routes into the suburbs to reap the rewards of land speculation. These real estate dealings represented the first perverse effects of regulating urban transportation in the United States.

---

[10] Alfred E. Kahn, *The Economics of Regulation: Principles and Institutions*, vol. 2 (New York: John Wiley and Sons, 1971).

[11] Ronald Kirby et al., *Para-transit: Neglected Options for Urban Mobility* (Washington, D.C.: The Urban Institute, 1974).

[12] See Sam Bass Warner, *Streetcar Suburbs: The Process of Growth in Boston* (Cambridge, Mass.: Harvard University Press, 1962); and Spencer Crump, *Ride the Big Red Cars: How Trolleys Built Southern California* (Los Angeles: Trans-Anglo Books, 1962).

By 1915 over 62,000 privately owned and operated automobile passenger services, called jitneys, were plying their trade in U.S. cities.[13] Within three years, however, nearly all of them were regulated out of existence by public authorities at the insistence of streetcar and trolley line owners who felt the jitneys were a threat to their businesses.[14] Taxis, which exploded in numbers during the Depression years when thousands of Americans were out of work, came under similar attack. Streetcar operators and later bus companies sought to restrict the entry of taxis in most U.S. cities to lessen their competition. Protestations that free competition would lead to violent taxi wars were common. Large taxi companies generally welcomed regulation in hopes of limiting the number of part-time drivers and of increasing fares.[15]

Although municipal authorities have historically regulated taxis, both state governments and the federal government have had a hand in suppressing them as well. The subsidization of transit fares has undermined the ability of taxis to compete in many urban transportation markets. Moreover, legislation that protects transit employees from being hurt by any program involving state and federal grants has retarded the emergence of taxis and other modal alternatives by ensuring continued high levels of transit operations.

Other services, such as private dial-a-ride vans and commuter club buses, have similarly been inhibited by transit authorities, which enjoy monopoly rights as sole service providers in their communities. In some states, carpools operated on a profit basis and large privately run vanpools come under restrictive regulations as well. The attitude of protectionism has become so pervasive in many urban transportation settings that free competition has all but vanished.

A byzantine network of local, regional, and state authorities has evolved for administering and enforcing these regulations. Often, local police departments and public utility or service commissions are empowered to regulate and license taxis, dial-a-ride services, and (in the few cities where they remain) jitneys. These authorities usually exercise control over fare and service features (for example, routes, hours of service, and fare structure) while safety matters typically remain vested in the state (for example, the licensing of drivers). In many cases regional transit authorities also exert control over the routing of taxis and dial-a-ride services within their district boundaries. In most metropolitan areas, a state regulatory commission has authority to oversee private common-carrier bus operations and vanpools. Where services cross state boundaries, the Interstate Commerce Commission has jurisdiction.

---

[13] R. D. Eckert and G. W. Hilton, "The Jitneys," *Journal of Law and Economics* 15 (October 1972): 293–325.

[14] Ibid., p. 294.

[15] Ibid., p. 295.

Usually, private bus companies must obtain a certificate of public convenience and necessity from the regulatory body before they initiate service. If unfair competition is alleged, the new entrant shoulders the burden of proof that the overall quality of service in the region will materially improve. Local and regional transit, taxi, and intercity bus competitors are generally involved in the review of service applications, and any one of the reviewers can often complain loudly enough to block entry. State regulatory commissions also usually place limits on school bus usage, restrict special chartered services in urban areas, and impose limits on the passenger loads of privately sponsored vanpools.

Regulatory agencies have generally developed a reputation for being conservative and favoring the status quo. The charge of these regulatory commissions has been to protect monopolistic rights of public transit systems, and most commissions tend to rigidly enforce existing laws in this pursuit. Existing services almost always prevail in unfair competition suits. Regulatory agencies have become identified as enforcers rather than advocates of change, as have most policing authorities. Regardless of the merits, the fact remains that this posture has impeded innovation and has inhibited the emergence of integrated urban transportation services. In view of the current situation, it is helpful to explore specific areas for regulatory reform.

Deregulating Taxi, Dial-a-Ride, and Jitney Services

Taxis, dial-a-ride vans, and jitneys, often referred to as forms of paratransit, have service features that fall between those of the private automobile and conventional bus transit. All three forms respond immediately to travel requests made by phone or curbside hail, and they often charge a premium fare for this feature. By comparison, other forms of paratransit, such as vanpools and commuter buses, provide services that are typically prearranged and operated on a fixed schedule between given points.

Taxis

The regulation of taxis in U.S. cities is universal, and few people question the propriety of regulating driver qualifications and insurance coverage. Limits on the number of taxis that can operate in a city, however, as well as limits on the types of services provided and the rates that can be charged have not generally been accepted by the traveling public. Most cities place a ceiling on the number of licenses (medallions) granted, often based on the number of cabs per capita. In the case of a handful of cities, notably Los Angeles and Chicago, exclusive franchises are granted to one company or a few companies. Because of these restrictive practices, large fleets offering fairly uniform-quality services have become the norm in most big cities.

Limits on supply more often than not have also meant relatively high fare rates. For instance, Washington, D.C., allows almost unrestricted entry into the taxi market and boasts a ratio of over 13 cabs per 1,000 residents, by far the highest in the country. For a typical four-mile trip, the cab fare in the nation's capital is only about $2.75. By comparison, New York City's notoriously restrictive taxi regulations have resulted in fewer than 1.5 cabs per 1,000 residents. A four-mile taxi ride in Manhattan costs about $4.75. With a New York City taxi medallion today bringing in $65,000 or more, it is clear that the cost of monopoly privileges are being passed on to customers.

Studies have found that travelers are more sensitive to the availability of taxis than they are to travel times, speeds, or almost any other service features.[16] Where taxis are given unrestricted freedom to ply their trade, the quality of urban transportation has generally improved. Not only do unregulated cities have more than three times as many cabs per capita, but services are often closely integrated with local bus and rail services as well.[17] Taxis have also proven their strength in low-density areas where public transit is highly unprofitable or uncompetitive. Where individual owner-operators are allowed to act as entrepreneurs, marginal markets abandoned by large fleets are again being served.

---

[16] Kirby et al., chap. 7.

[17] Mark Frankena and Paul Paulter, *An Economic Analysis of Taxicab Regulation* (Washington, D.C.: Bureau of Economics, Federal Trade Commission, 1984).

Deregulation would be a particular boon to small taxi companies and private individuals who are currently denied entrepreneurial freedom. Moreover, lifting entry controls should be expected to increase employment opportunities for some urban residents, particularly among low-income and minority populations in which joblessness is the highest.[18]

Shared-Ride and Dial-a-Ride Services

Another constructive regulatory reform would be to allow cabs to pick up more than one party en route. The serving of multiple passengers by cab is often referred to as shared-ride taxi; when a van-sized vehicle is dispatched to pick up and unload unrelated trips with only minor route deviations, the service is referred to as dial-a-ride. Dial-a-ride vans are often run by private, profitmaking companies that receive public subsidies for serving special passenger groups, such as the elderly and physically disabled. Shared-ride taxis, on the other hand, are typically unsubsidized.

During World War II, shared-ride taxis flourished in Washington, D.C., and they have had a modest resurgence there since the 1970s.[19] Washington taxi drivers openly displayed destination signs on their front windows during the war period, allowing people to hail cabs going their direction. Riders received a break in fares, and scarce war-time resources were being used efficiently. In 1974 Washington again adopted a version of taxi ride-sharing, primarily in response to gasoline shortages.

As in the case of the nation's capital, use of the shared-ride concept in other areas would require the decontrol of taxi price structures to allow zonal fares rather than distance metering so that riders would not be overcharged. It would not be necessary to operate all cabs on a shared-ride basis, however, as some passengers could be dissuaded by even modest time delays in picking up other fares. A mix of exclusive-ride and shared-ride taxis, however, is the best way to satisfy the riding public's preferences.

Experiences with shared-ride taxis in El Cajon, Calif., Davenport, Iowa, and some 20 other cities have been positive.[20] Shared-ride taxis have operated at a much lower cost per passenger than subsidized dial-a-ride vans. Much of the cost savings has been attributed to lower wage levels and higher productivity among privately employed taxi workers. Shared-ride taxis have actually been a boon to local bus systems in these areas by siphoning off some of the already oversaturated peak demands, in addition to serving as feeders into bus lines and rail stations. They have proven particularly cost-effective in suburban, peak-hour markets.

---

[18] Patricia Buckley, "The Deregulation of Urban Taxicab Markets: A Note," *American Economist* 27 (Spring 1982): 73–76.

[19] Kirby et al., chap. 7.

[20] Roger Teal et al., "Subsidized Shared-Ride Taxi Services," *Transportation Research Record* 778 (1980): 25–32.

The shedding of peak loads to private taxi operators could result in real cost-savings to public transit. The peak is transit's nemesis, largely because of restrictive work rules requiring time-and-half pay for spreading drivers' duties over the morning and evening peaks. Studies have shown the cost of running an additional bus during rush hours to be two to three times as much as during the off-peak period.[21] The shedding of peak loads to shared-ride taxis would likely be welcomed by transit managers, as well as by commuters seeking a service of a higher quality than that offered by buses.

U.S. experiences with substituting shared-ride taxis for fixed-route bus services on a contract basis have proven particularly successful. In Phoenix, for example, the local transit authority contracted with a taxi company to replace minimal-level Sunday fixed-route bus services—an arrangement that resulted in an estimated cost savings of $600,000 per year.[22] Similar arrangements have recently been made for evening services in Ann Arbor, Mich., and Chapel Hill, N.C. In the Norfolk–Chesapeake–Virginia Beach area, a number of poorly patronized bus runs have been replaced with private services that use smaller vehicles tailored to lower demand levels; taxis operate between suburban residential neighborhoods and a number of shopping–office complexes in the region, replacing comparable bus services at an estimated savings of $16 per hour.[23]

Another advantage of incorporating shared-ride taxis into the urban transportation market is that they can also provide more specialized curb-to-curb services for elderly and handicapped populations.[24] Programs that have given senior citizens, disabled persons, and poor people travel vouchers (that is, user-side subsidies) for riding their choice of transportation—be it bus, shared-ride taxi, dial-up van service—have encouraged healthy competition among different service providers while serving an important public need. Clearly, shared-ride taxis would mesh nicely into most cities' transportation milieu by complementing peak-hour transit operations and providing a necessary off-peak service to disadvantaged persons.

---

[21] See J. Boyd et al., *Evaluation of Rail Rapid Transit and Express Bus Service in Urban Commuter Markets* (Washington, D.C.: Institute of the Defense, 1973); Charles Lee and I. Steedman, "Economies of Scale in Bus Transportation," *Journal of Transport Economics and Policy* 4 (January 1970): 15–28; and J. Stuart Wabe and Oliver B. Coles, "The Peak and Off-Peak Demand for Bus Transport: A Cross-sectional Analysis of British Municipal Operators," *Applied Economics* 7 (March 1975): 25–30.

[22] Ronald Fisher, "Megatrends in Urban Transport," *Transportation Quarterly* 38 (January 1984): 87–102.

[23] A. Jeff Becker and James C. Echols, "Paratransit at a Transit Agency: The Experience in Norfolk, Virginia," *Transportation Research Record* 914 (1983): 49–57.

[24] Roger Teal et al., *Taxi-Based Special Transit Services* (Washington, D.C.: Urban Mass Transportation Administration, U.S. Department of Transportation, 1983).

Jitneys

Jitneys extend the shared-ride concept even further. Perhaps the purest form of transportation entrepreneurship, jitneys (cars or station wagons with room for 5–12 passengers) pick up and drop off patrons over a semi-fixed route on a fairly regular basis. Customers are generally taken until the vehicle is full, and only slight detours from a major street are typically made. Political pressure mounted by trolley operators to bar jitneys from "skimming the cream" resulted in the banning of jitneys in most cities. Although jitneys may well have threatened municipal transit systems in 1915, a time when those systems were in their infancy and struggling to survive, they would actually benefit urban transit systems today by providing, as shared-ride taxis do, a much-needed supplement to peak-period capacity.

Publicly sanctioned jitneys currently operate in San Francisco, Atlantic City, and, most recently, San Diego, but ordinances restrict their numbers to below 500. By contrast, in many Latin American, Asian, and Middle East cities, jitneys are the chief mode of urban transportation. It has been estimated, for instance, that in Caracas, Buenos Aires, and Istanbul jitneys serve over half of the daily commuters.[25]

Despite restrictive ordinances, the market demand for jitneys has become so great that they operate illegally in Chicago, Pittsburgh, Baton Rouge, Miami, Chattanooga, and probably other cities as well. In Chattanooga alone, it has been estimated that over 85 jitneys illegally serve over 20 million willing customers a year.[26] These clandestine operations generally thrive in low-income, minority communities where a market exists for a service that is a hybrid between more expensive taxis and less convenient bus transit. Authorities have tended to look the other way when confronting these illegal, albeit successful, operations.

---

[25] Robert Poole, "Alternatives to Government-Financed Public Transit," *Oakland Tribune*, 30 September 1980.

[26] Ibid.

Recent Experiments with Taxi and Jitney Deregulation

Recent experiences with taxi entry and fare deregulation in 22 U.S. cities over the past five years, and with jitney deregulation in San Diego, have been encouraging.[27] Most areas either have removed the ceiling on taxi permits or have raised it significantly. Some have permitted ride-sharing in taxis, along with the introduction of zonal fares. Some cities also have allowed exclusive-ride fares to vary. Although taxi operators and transit interests have fought these open-door policies in most places, strong city council and public support for these programs ultimately have prevailed.

In almost every setting, the number of firms and cab service hours have increased markedly since deregulation. Between 1979 and 1983, for instance, the total number of cab permits in Seattle rose 178 percent. Small cab companies and private owner-operators have proliferated the most. In Seattle, for instance, small fleets (those with four to 13 cabs) increased in number from nine to 23, whereas the share of cabs held by the three largest firms declined from 70 percent to 54 percent.

More cabs have generally meant more and better service (in particular, shorter waits, fewer nonresponses to phone requests, and cleaner vehicles). Total weekly hours of cab service in San Diego, for example, has increased 26 percent since deregulation. Passenger waits at major cabstands have almost disappeared in several places. Average waits for San Diego's radio-dispatched cabs, moreover, fell from 10 minutes to 8 minutes in the first two years of deregulation. In Seattle, price decontrol has also led to a variety of fare structures, including off-peak discounts and lower fares for repeat, advanced-reservation customers.

Overall, fares have essentially remained unchanged (in real terms) since deregulation in most cities, while service quality has generally improved (in particular, shorter waits, fewer nonresponses to phone requests, and cleaner vehicles). In addition, decontrol has also led to greater market specialization, with smaller and newer operators concentrating on hail and long-haul business, and the larger and older companies going after the phone-request and package-delivery business.

San Diego's experiences with legalized jitneys have been equally impressive. By early 1983 15 jitney companies, owning a total of 48 licensed vehicles and serving nearly 12,000 weekly customers, had entered the market. They operate on streets paralleling the new light-rail trolley and main bus routes, concentrating mainly on commercial strips and tourist spots. Jitney and shared-ride taxi fares can be set at any level as long as they are posted in the vehicle's front window (in two-inch-high lettering) and do not exceed a maximum rate. Fares have proven to be a real bargain. A five-mile trip from San Diego's airport to the downtown area, for instance, today costs about $3 by jitney compared to $12 by exclusive-ride taxi.

---

[27] See Frankena and Paulter; and Pat Gelb, *Effects of Taxi Regulatory Revision in Seattle, Washington, Effects of Taxi Regulatory Revision in San Diego, California,* and *Taxis Regulatory Revision in Portland, Oregon: A Case Study* (Washington, D.C.: U.S. Department of Transportation, Urban Mass Transportation Administration, 1983, 1984).

The only major snag with deregulation to date in most cities has been isolated incidences of price-gouging, particularly at airports, where there are tourists unfamiliar with the competitive pricing program. Seattle, San Diego, and several other cities have subsequently imposed fare ceilings on airport cabs. San Diego, moreover, placed a moratorium on new cab permits in 1983 to give the city council time to reassess the entire program. Overall, however, residents and visitors in all three cities have materially benefited from the specialization of services and revised pricing made possible by deregulation.

Deregulating Vanpools and Commuter Buses

Vanpools and commuter buses—paratransit modes that provide prearranged or book-in-advance commuting—have similarly been stifled by regulation, although perhaps to a lesser extent than taxis and jitneys. Court rulings in a number of states during the 1970s generally held that vanpools are public carriers and thus subject to various certification requirements. Some courts even interpreted vanpools to be illegal bus lines. Several states, notably California and Tennessee, subsequently passed legislation exempting employer-sponsored vehicles carrying 15 or fewer passengers from any government regulations, while in others (such as Washington, D.C.) fairly strict compliance requirements remained in place.[28] Voluntary, share-the-expense carpools are permitted in all states, while nonemployment-related van services operating on a profit basis are generally prohibited

Commuter buses have become particularly popular in recent years, providing subscribers with high-quality services (for example, comfortable seats and express rides) between suburban neighborhoods and central office locations. Today, subscription buses commute regularly to government offices in Washington, D.C., San Francisco's financial district, and military bases in Virginia's tidewater area. The largest fleets of subscription buses in the United States, however, are in Los Angeles and Chicago, where over 5,000 commuters ride in comfort to major employment centers in each area.

---

[28] Roger Teal and Genevieve Giuliano, "Increasing the Role of the Private Sector in Commuter Bus Service Provision," *Built Environment* 8 (1982): 172–83.

Although there appears to be a much larger market demand for subscription services, local transit authorities have succeeded in holding down the number of subscription services. A recent study by the Southern California Association of Governments estimates that 22 public transit routes in the area could be replaced by subscription buses for about one-half of current costs, saving over $5 million annually.[29] Yet transit authorities have filed protests against a number of commuter bus operations. The reactions of transit properties toward private competition in other parts of the country, however, have been much different. In Marin County, Calif., for example, the regional transit authority was instrumental in establishing a number of private commuter van and bus services simply because it realized that the cost of running its own buses along certain corridors would be prohibitive.

Most recently Tennessee lifted controls over private bus operations. The state Public Service Commission has designated certain counties as "citizen transportation areas" to allow privately owned vehicles to be used for passenger services. Church and other special-purpose buses are doubling as commuter vehicles in these counties. These reforms have provided vital travel alternatives to residents of small communities that have recently lost intercity bus services.

Deregulating Parking

Relaxing the parking requirements of municipal zoning ordinances should also be a reformative priority. Most ordinances mandate a minimum number of off-street parking spaces based on the intensity of land use—for example, per dwelling unit or per square foot of office space. Although minimum-parking-space regulations aim to ensure that incoming traffic can be adequately handled, more often than not they also serve to establish a strong automobile orientation along highway corridors. Repeated practices of requiring a minimum number of parking spaces have unquestionably been at the expense of relegating public transit to more of a second-class status.

---

[29] Southern California Association of Governments, "Commuter and Express Bus Service in the SCAG Region: A Policy Analysis of Public and Private Operations," unpublished report (Los Angeles, 1982).

Minimum-parking-space regulations, in some sense, become self-fulfilling prophecies. They are based on the assumption that parking-space requirements reflect a "need" to travel by automobile and that consequently they encourage private vehicle use. In addition, there is a rule-of-thumb air about such regulations that is particularly disquieting. One study, for instance, found that minimum parking space requirements for a 10,000-square-foot office building in California varied from 10 in Long Beach to 80 in Placentia.[30] Moreover, local planning departments often bargain for more than the minimum number of spaces in negotiating with private developers. The fact that some ordinances require more parking than what the private market will bear while others require less points to the general arbitrariness of these regulations.[31] Both options are reasonable—congestion can be reduced by either getting cars off the streets once they reach their destination or by limiting the number of vehicles that can park at a site. The best way to balance these competing rationales, however, is to let market forces determine the appropriate level of parking supply.

Zoning regulations effectively reduce the price of parking and encourage long-term parking. Congestion is often aggravated in that automobile trips converge on a single area. Minimum-parking-space regulations also distort the land market because they effectively become a tax on the quantity of floor space in a new building. Where more spaces are required beyond that dictated by the market, inner-city redevelopment can be discouraged. Also, the parking tax can shift new development away from downtown areas well served by public transit to lower-density areas where land can be acquired more cheaply for parking construction.

Rigid parking regulations have generally been unresponsive to changing parking needs brought on by rising fuel costs and the switch to smaller vehicles. For these very reasons, according to the Urban Land Institute, most shopping center developers are now willing to provide up to 18 percent fewer spaces than in years past.[32] Some cities have recently modified their zoning ordinances in recognition of changing characteristics of demand and broader transportation objectives.

Recently, Seattle, San Francisco, and Portland, Ore., have all lifted minimum-parking-space requirements for new downtown developments, while Chicago and Boston have placed a ban on new downtown parking construction altogether.[33] Since the late 1970s Seattle has witnessed a 2 percent reduction in total parking supply, even though 13 major proposals for new development have been introduced. Other cities now allow private developers to make "in-lieu" cash contributions to support transit or ridesharing programs for their tenants in exchange for reducing or eliminating off-street parking requirements. In both Los Angeles and Palo Alto, Calif., builders provide "effective alternatives to auto access," such as vanpool leasing and cash payments to transit agencies in return for less-stringent parking requirements. Fee-in-lieu-of-parking programs also exist in Toronto, Chicago, and New York for the purpose of pooling monies for building central parking garages and supporting local transit services. And in Montgomery County, Md., 5–35 percent reductions in the number of parking spaces are allowed for new buildings in close proximity to Metrorail stations.[34]

---

[30] Donald Shoup and Don Pickrell, "Problems with Parking Requirements in Zoning Ordinances," *Traffic Quarterly* 32 (July 1978): 545–61.

[31] Ibid., p. 551.

[32] Urban Land Institute, *Parking Requirements for Shopping Centers: Summary Recommendations and Research Study Report* (Washington, D.C., 1982).

The easing of parking regulations would probably be one of the more effective means of modifying current travel behavior. A stronger reliance on market pressures to set the level of parking supply would undoubtedly benefit public transit and urban redevelopment efforts.

Other Candidates for Deregulation

Other regulations have hampered progress in the urban transportation sector. The subdivision regulations of most U.S. communities place a lid on the allowable densities of various types of land uses. Maximum density requirements are usually set to comply with regional development goals and to ensure that excessive demands will not be placed on local sewer, water, and road facilities. In more suburban settings, regulations have generally encouraged the provision of ample roadway capacity and off-street parking, often as a precondition to the approval of a subdivision construction request. Local planners have also tended to recommend the downsizing of new subdivisions whenever traffic impact analyses suggest that the additional automobile travel will congest nearby intersections.

The problem with this low-density bias is that it precludes the development of the customer base needed to support transit and ridesharing alternatives to the automobile. Subdivision regulations are generally myopic regarding transportation matters; they are more concerned with immediate traffic impact issues rather than with building a longer-term mass-transit orientation. Modifying subdivision regulations to allow some flexibility in the establishment of residential densities could prove to be a tremendous inducement to public transit and other forms of ridesharing over the long run.

---

[33] Wilbur Smith, "What's New in Parking," *Planning* (June 1983): 10–14.

[34] Ibid., p. 13.

Repressive controls over urban transportation have not come solely from the local level. The federal government, in particular, has had a direct hand in the financial problems that have plagued urban transit systems by imposing various requirements (often tied to subsidy programs) that effectively increase the costs of local transit services. Among these are mandates involving labor protection and fully accessible transit vehicles. Section 13(c) of the amended Urban Mass Transportation Act of 1964 guarantees that transit employees will not be adversely affected by any program involving federal transit grants.[35] This stipulation has been blamed for tying the hands of transit management during contract negotiations by giving labor the equivalent of veto power over federal grants. In addition to augmenting the resource base available for generous settlements, 13(c) has also promoted work rules involving guaranteed pay clauses and prohibitions on part-time employment. Moreover, 13(c) has prevented taxis from providing lower-cost services along many urban corridors. In Norfolk, Va., for instance, unions used 13(c) in suing the local transit agency when unproductive routes were turned over to a private shared-ride taxi operator.[36] Similarly, the Davis-Bacon Act, which requires that the highest prevailing union wage level be paid on federally funded construction projects, has also been cited as a cost escalator—particularly on new rapid rail systems in Washington, D.C., and Atlanta. In that labor expenses account for roughly 70 percent of the cost of operating most transit systems in the United States, the fiscal consequences of these regulations have been substantial.

Federal regulations mandating barrier-free accessibility and off-peak fare discounts for elderly and handicapped persons have also been criticized for their adverse financial effects. Equal accessibility requirements place substantial capital cost burdens on local operators, while special pricing mandates suppress farebox income. The Congressional Budget Office has projected, for instance, that the ultimate cost of providing fully accessible buses and rail stations to handicapped passengers would reach $38 per trip.[37] By comparison, demand-responsive taxis would serve about 3.5 times the number of severely disabled persons at one-fifth the cost. Although the Reagan administration has loosened some of the full-accessibility requirements, the combined introduction of user-side subsidies and deregulation of shared-ride services would do much more for America's elderly and handicapped than these public mandates.

---

[35] Alan Altshuler, *Testimony before the Subcommittee on Investigation and Oversight, Committee on Public Works and Transportation,* unpublished testimony (Washington, D.C.: U.S. House of Representatives, 23 June 1981).

[36] Becker and Echols.

Conclusions

Regulations governing urban transportation have been built up, layer by layer, over time to the point where today they represent major obstacles to innovation. Foremost, they have impeded the emergence of jitneys, vanpools, shared-ride taxis, private club buses, school bus charter services, and a host of other alternatives to the automobile. Repealing present-day controls to allow freer entry would probably do more to alleviate urban transportation problems than would any assortment of technological fixes.

The removal of entry restrictions and exclusive franchise privileges would open the way for a rich mix of new services to penetrate urban transportation markets. Fare deregulation, moreover, would enable shared-ride services to be offered and would likely lead to time-of-day charges to reflect the higher costs of peak period services. Even in Washington, D.C., where taxis are fairly plentiful, there is still a shortage of cabs during rush hours. Experiences in San Diego, Seattle, and other places are graphic testimonies to the potential benefits of paratransit deregulation.

Perhaps most disconcerting has been the general arbitrariness of many regulations to date. Jitneys have been practically regulated out of existence, whereas carpools have been almost untouched by regulation. In some localities, zoning ordinances require abundant off-street parking while in neighboring communities only minimal on-site parking may be called for. Some states place stringent controls on subscription bus services while others actively encourage this form of ride sharing. Such inconsistencies cast doubt on the underlying rationales behind these controls and suggest that private market forces would probably be more effective in allocating resources within the urban transportation sector.

Deregulation would not be without side effects, however. For one, certain groups could suffer particularly if public transit services were replaced by paratransit on a wholesale scale. A shift toward higher-quality services priced at premium fares would increase travel options for more affluent commuters while perhaps diminishing them for poorer persons. It is also possible that some carriers would pursue exclusionary practices, such as refusing to serve minority neighborhoods. To date, however, there has been no incidence of such discrimination among deregulated cities. Moreover, the fare reductions that result from taxi decontrol would likely benefit the poor the most since they spend a larger share of their incomes on taxis than other groups. User-side subsidies could also be introduced to help defray the travel costs of low-income residents. Overall, it would seem more likely that a deregulated environment would offer all Americans a richer assortment of travel options than they now have.

---

[37] *Urban Transportation for Handicapped Persons: Alternative Federal Approaches* (Washington, D.C.: U.S. Congressional Budget Office, 1979).

The relaxation of entry restrictions on taxis and shared-ride services would also pose some inequities among cab owners. Those who have paid as much as $65,000 for taxi medallions, for instance, would stand to lose a lot of money from the decontrol of taxi entry. A municipality could opt to buy back all medallions at their purchase price, but not without a substantial cash outlay. On the other hand, it may also be argued that those who benefited from regulation have no implicit right to special protection and must bear the risks of deregulation—no one promised them protection in perpetuity. Still, those who paid high entry fees under the regulatory regime could be the first to file suit against localities to recoup their losses.

The decontrol of entry regulations may also be expected to undermine transit services along certain corridors where a single service provider could most economically operate. Given the diversity of travel demands in the United States, however, areas where a true natural monopoly could be justified would likely be the exception rather than the rule. Past research suggests that with the exception of rapid rail operations, most transit services operate under conditions of constant returns to scale.[38] Thus the natural monopoly argument for regulating entry into the urban transportation market would seem to have limited application. On balance the benefits accruing to the traveling public from deregulation would far offset any detrimental effects a few transit systems may experience from heavy competition.

In closing, deregulation would not give rise to a purely laissez-faire system of urban transportation, nor would it produce the chaos and disregard for traffic laws found in some highly congested Third World cities where private transportation flourishes. Clearly, controls over safety, driver qualifications, and operating practices would still be in order. However, there seems to be no compelling reason why price and supply controls should be imposed on taxis, jitneys, vanpools, club buses, and parking. The marketplace has proven its prowess at responding to the many and varied preferences of American consumers in other areas, so there is no reason why the same would not hold true in the urban transportation sector. What is needed more than ever in our cities is a freely competitive transportation environment—one in which services and prices can be closely tailored to the diverse needs of America's traveling public.

---

[38] Joseph Berechman, "Analysis of Costs, Economies of Scale and Factor Demand in Bus Transport," working paper (Irvine: University of California, Institute of Transportation Studies, 1982).

JOB PROTECTION IN URBAN MASS TRANSIT

Simon Rottenberg

Introduction

The urban mass transit industry in the United States has been a declining industry, especially since the Second World War, despite massive subsidies for the acquisition of capital equipment and facilities and for the operating expenses of the system.[1] Rising real income that caused the ownership of automobiles to become more common, the improvement of the highway capital stock, and the growth of suburban residential areas, coupled with the qualities of accessibility and convenience, caused the privately owned automobile to become the overwhelmingly preferred mode of transportation to and from work and for shopping and leisure travel within cities.

Before World War II, urban mass transit service was mainly provided by privately owned companies that held monopoly franchises and were subject to the regulation of the public authorities with respect to service quality and rates. With the spread of the automobile as a competitive mode of transit, it became clear that the socially efficient response would be the shrinkage of the urban mass transit industry, the withdrawal of monopoly franchises, and the free entry of paratransit modes into the interstices of the system where demand

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Professor of Economics at the University of Massachusetts.

[1] For the general treatment of the urban mass transit system and its subsidization, see John R. Meyer and José A. Gómez-Ibáñez, *Autos, Transit and Cities* (Cambridge, Mass.: Harvard University Press, 1981); Alan Altshuler (with James P. Womack and John R. Pucher), *The Urban Transportation System, Politics and Policy Innovation* (Cambridge, Mass.: MIT Press, 1979); George W. Hilton, *Federal Transit Subsidies* (Washington, D.C.: American Enterprise Institute, 1974); Wilfred Owen, *The Metropolitan Transportation Problem* (Washington, D.C.: The Brookings Institution, 1966); John R. Meyer, John F. Kain, and Martin Wohl, *The Urban Transportation Problem* (Cambridge, Mass.: Harvard University Press, 1965).

was sufficiently intense to yield a normal return on the provision of transit services.

This process was begun in the mid-1940s and 1950s with the withdrawal from the industry of private companies that were encountering operating losses. But this shakeout came to an end when the Congress, in 1964, enacted the Urban Mass Transportation Act that provided federal funds for urban transit capital acquisitions. State and local public agencies were established to use federal grants for the acquisition of the capital assets of the withdrawing private companies and the monopoly franchises were transferred to the new public sector agencies.

In 1940 there were only 20 publicly owned urban transit systems; in 1982 there were 581.[2] In 1940 publicly owned systems owned and leased only 7 percent of all the vehicles used by the industry; in 1982 this had become 91 percent. In 1945 there were 23.2 billion passenger rides; in 1980 this had shrunk to 8.2 billion. Despite the diminution in passenger load, the number of vehicles employed by the industry has remained much more constant between the beginning and end of this period; in 1945 the industry used 75 thousand vehicles and in 1982, 74 thousand, with light rail vehicles being displaced by motor buses.

The urban transit industry suffers immense losses each year. Its passenger fare and other operating revenues in 1982 covered only 40 percent of its operating expenses. To cover operating losses, the system received in 1982, $4.6 billion in subsidies—about 80 percent from local and state governments and the remainder from federal operating grants. Since 1975 the federal government has provided almost $6 billion in grants to cover operating deficits. In addition to operating subsidies, the system is also heavily subsidized for capital acquisitions and improvements. Since 1965 the federal government has provided over $20 billion to the industry in capital grants.

Given that urban mass transit systems produce transport services for local users in the communities in which they operate, it is questionable whether the subsidization of those systems by the federal government is warranted. The subsidies imply massive income transfers from nonusers to users of the transit systems and from residents in other parts of the country to residents in the large urban concentrations.

---

[2] These and other data are derived from the American Public Transit Association, *Transit Fact Book, 1981* (updated by APTA telephone interview) and John R. Meyer and José A. Gómez-Ibáñez, *Autos, Transit and Cities.*

A goodly share of the subsidies have been siphoned off by employees of the urban transit systems who, through their unions, have exercised monopoly power in the urban transit labor markets and, thus, gained economic rents for themselves. Although the number of passenger trips has been reduced from 23 billion in 1945 to 8 billion in 1980, the number of transit system employees has fallen during the same period, from 242,000 to 189,000. Despite widely fluctuating demand for transit service over the hours of the day, which would seem to warrant the intensive employment of part-time workers, the unions of the industry raise obstacles to the employment of part-time workers, so that, in 1980, there were only 4,600 part-time employees out of a total of 189,000 employees. While the skills necessary for operating transit vehicles are widely diffused in the country's adult population—so that one would expect compensation rates in the industry to be relatively low—the exercise of union monopoly power, reinforced by the availability of government subsidies and by the existence of monopoly power in the system's product market, has caused earnings of employees in the industry to be similar to those of persons with advanced graduate schooling. It has been estimated that, from 1970 to 1980, 33 percent of federal government operating subsidies in urban transit were siphoned off by employees in the form of increased compensation; 21 percent of the subsidies were consumed by increased utilization of labor, as a result of work rules enforced by the unions and other inefficiencies generated by subsidies; and only 27 percent of the subsidies were reflected in benefits to transit system users in the form of fare reductions and additional trips.[3]

Monopoly on the selling side of the urban transit labor market is entrenched and reinforced by section 13(c) of the Urban Mass Transportation Act of 1964. That act, as amended, authorizes the Secretary of Transportation to make grants to urban transit systems for capital acquisitions and modifications, for transit systems operations, and for other purposes related to the "development of comprehensive and coordinated mass transportation systems, both public and private, in metropolitan and other urban areas."

Section 13(c) of that act makes provision for the protection of workers who may be affected. It reads:

---

[3] Douglass B. Lee, *Evaluation of Federal Operating Subsidies to Transit,* Staff Study SS-67-U.3.03, U.S. Department of Transportation, Transportation Systems Center, Cambridge, Mass., July 1983, duplicated. The remainder of federal operating subsidies not reported above were used to cover price increases for fuel and materials, fuel consumption, and for the provision of unutilized service.

> It shall be a condition of any assistance under section 3 of this Act that fair and equitable arrangements are made, as determined by the Secretary of Labor, to protect the interests of employees affected by such assistance. Such protective arrangements shall include, without being limited to, such provisions as may be necessary for (1) the preservation of rights, privileges, and benefits (including continuation of pension rights and benefits) under existing collective bargaining agreements or otherwise; (2) the continuation of collective bargaining rights; (3) the protection of individual employees against a worsening of their positions with respect to their employment; (4) assurances of employment to employees of acquired mass transportation systems and priority or reemployment of employees terminated or laid off; and (5) paid training or retraining programs. Such arrangements shall include provisions protecting individual employees against a worsening of their positions with respect to their employment which shall in no event provide benefits less than those established pursuant to section 5(2) (f) of the (Interstate Commerce Act of 1877) (which provides that for a period of four years employees who are affected will not be in a worse position with respect to their employment.) The contract for the granting of any such assistance shall specify the terms and conditions of the protective arrangement.

Politicization of Mass Transit

Since 1964 urban mass transportation essentially has become an exclusively public sector activity; a large number of private firms offering transit services survive, but they carry a very small proportion of all passengers. When the industry was privately owned, it was based on monopoly franchises awarded by the public authorities; its fares and the level of services it was required to offer were regulated. As a public sector activity, fares and service levels are responsive to pressures emanating from political sources. In both the private franchise and the ensuing public-ownership services, the mass transportation system has operated essentially as a monopoly in the transit product market.

The transit industry has been declining since the end of World War II, although the number of passengers carried stabilized and rose slightly in the 1970s. Apparently in response to rising incomes, the construction of road networks, and the diffusion of population within urban areas, the automobile has become the primary mode of passenger transport, and it appears that it will continue to perform this role. Mass transit is numerically important only in a handful of large, densely populated metropolitan areas.

The number of passengers carried by conventional mass transit vehicles varies greatly among routes and across hours of the day; the mass transit industry is faced with demand peaks and troughs when seen in the contexts of both time of day and routes served. Within this context, the industry has experienced ever-enlarging operating deficits since about 1965. Operating revenues have risen slightly, but operating expenses have risen at a much more rapid rate since the late 1960s. The deficits have been covered by enlarged subsidies from federal, state, and local sources. Mean fares have diminished in real terms in the 1970s. Capital costs also have been covered, almost completely, by public subsidies.

The mass transit industry is labor-intensive. A very large fraction of operating costs are labor costs, paid partly in cash wages and partly as fringes of various kinds. Of $6 billion in transit operating costs in 1980, 73 percent was expended for salaries, wages, and fringe benefits.[4] It has long been characteristic of the industry that its workers have been represented by unions in the joint determination of rules of work and levels of compensation. That pattern, established when the industry was privately owned, was prolonged into the period of public ownership by the labor protection provisions of the Urban Mass Transportation Act of 1964. By the exercise of monopoly power in the urban transit labor market, the unions have been able to produce rents for the industry's employees. These rents take the form of earnings that appear to be greatly in excess of those that would be necessary in a competitive labor market to attract the relevant number of workers to accept employment in the industry. As noted, the skills for the operation of motor vehicles are commonly possessed by almost all adults in the country's population.

Mass transit seems to have a comparative advantage over alternative modes of travel for commuting trips to and from work on densely traveled radial routes to central cities and for short trips within central cities that serve low-income neighborhoods. But conventional mass transit services are now grossly extended beyond those limits.

Many of the routes that are now served by large, regularly scheduled vehicles of the transit system appear to be more appropriate to service by paratransit vehicles of various kinds, including passenger-driven vans and car-pools. Paratransit service could also be fruitfully employed as a complement to conventional transit service on the radial and other routes at peak demand hours where conventional service is most productively employed. Thus, paratransit might offer both parallel and feeder services.

---

[4] American Public Transit Association, *Transit Fact Bood, 1981*, p. 48.

The shrinking of routes and neighborhoods that are served by conventional mass transit might well involve substantial reductions in the operating and maintenance personnel and, hence, the operating costs of the systems. But labor protection arrangements, designed to prevent adverse effects on workers or requiring the compensation of adversely affected workers to "make them whole," inhibit the kinds of adjustments that are necessary to achieve a socially optimal transit system. Such arrangements tend to perpetuate real resource inefficiencies in the organization and operation of urban mass transit systems.

Circumventing Section 13(c)

Grants to which section 13(c) protections are attached cause real and money costs in mass transit to be higher than they otherwise would be.[5] Either the services that the grants generate must be administered in ways that do not occasion the dismissal of transit workers (even if the services can be provided at lower real cost and with equal efficiency with fewer workers) or dismissed workers must be compensated for their dismissal, retrained, or relocated. Section 13(c) also requires collective bargaining in transit operations that are covered by federal subsidies, which reinforces the monopoly power that collective bargaining has produced in this industry.

The real and money costs of transit operations might be reduced if 13(c) costs could be avoided by making no (or fewer) operating subsidy grants to transit systems. If conventional transit systems were then to contract (as, for example, by a diminution of routes served), the gaps in the transit service pattern might be filled by a diverse set of alternative modes that would be free of 13(c) constraints. These include passenger-driven van and car-pools, employer-provided vanpool services, and, in densely traveled periods and routes, privately owned commercial transit operations that would charge fares sufficient to cover operating and capital expenses. These alternatives would be much cheaper than the existing capital-intensive strategies for reducing labor costs that are generated by 13(c) protections against adverse effects and 13(c) reinforcement of collective bargaining modes.

---

[5] Real costs occur when real resources used by the transit system (including labor) have alternative uses that society finds valuable. "Higher real costs" means that more resources are used than otherwise would be the case for the production of a given quantity of transit services. Money costs are money expenditures of the transit system. They may be used to pay for the resources the system employs, but they include, in addition, transfer payments, such as monopoly rents paid to transit employees that are *not* payments for labor services, and compensation payments made to dismissed workers.

Rent Seeking under 13(c)

It seems to be clear on its face that the money cost of operating the transit system is much enlarged because its employees are paid rents. Labor market rents are payments made in excess of those that would be necessary to induce the desired number of people to make their services available to a trade or occupation, provided there were competition in the labor market and no constraints on entry into or exit from the relevant occupation. Rents come into existence only when the labor market becomes less than fully competitive; they cannot exist in conditions of competition.

A very large fraction of the employees of the transit industry are motor vehicle operators. Successful performance in that occupation requires motor vehicle operating skill, a calm and courteous demeanor, good moral character, and capacity to read and to tell time. These are skills and personality traits that are commonly possessed by the country's adult population. Virtually every adult possesses some motor vehicle operating skills; only a short period of additional training would seem necessary to operate larger vehicles such as motor buses. In a competitive labor market the payments for the performance of vehicle operating services would be very low, given the ubiquity of the skill in the population. Under competition, the required number of operating personnel ought to be able to be attracted to the industry at very low wages.

The earnings of vehicle operating personnel are, however, much higher. Not only are base hourly wages higher than those that would occur under competition, but earnings are additionally enlarged by premium payment arrangements of various kinds, including overtime, spread-time, and shift premiums. The earnings include, additionally, payments for time, service, and status that do not represent vehicle operation time.[6]

In addition, the collective bargaining practice of the industry commonly links the earnings of employees in other occupational classes to the earnings of vehicle operators. Rents paid to operators, therefore, will be compounded by rents paid to nonoperator employees.

That a rent component is included in the payments made to transit employees is corroborated by the existence of the priority of reemployment provision in section 13(c) of the act and of seniority prescriptions in 13(c) agreements and in collective agreements that rank-order workers in the allocation of various kinds of opportunities. Such provisions are rationing devices. They determine who receives preference in a situation where there are more qualified applicants than there are jobs to be filled. They necessarily imply that earnings are in excess of those that would prevail under competition, since competitive markets clear and there is no role for nonprice rationing.

---

[6] In the idiomatic usage of the transit industry, vehicle operators are paid for report time, turn-in time, travel time between an operating station and a point of relief, intervening time, paid break time, minimum guarantees, student training time, standby time, time spent on union duties, and run-selection time.

Additional evidence of the existence of rents in the payments made to transit employees occurs in the form of relatively low voluntary separation rates in that industry. More attractive relative earnings opportunities in an industry will be associated with lower rates of voluntary quits in that industry. The payment of economic rents to labor will cause voluntary separations to be lower than in other industries where the payment for labor does not include a component of rent.

The rent component in the payments to transit labor is, of course, produced by the collective bargaining process. Section 13(c) affirms and entrenches collective bargaining in the industry and intensifies the monopoly power on the supply side of the transit labor market that produces the rents.

That section explicitly affirms in its text that "the continuation of collective bargaining rights" shall be a condition for the grant of assistance. It thus opens opportunities for the use of the 13(c) agreement as an instrument for securing, in the collective bargaining agreement, more favorable terms of employment than might otherwise be achieved. Furthermore, the grant of assistance under the Act is conditional on the signing of a 13(c) agreement by the relevant trade unions of the transit industry and by transit management.

Additionally, 13(c) agreements sometimes contain clauses that relate not to the forestalling of adverse effects upon employees as a result of transit assistance, but rather, have substantive implications themselves. A not unrepresentative clause provides, for example, that "in the event of *any* labor dispute," arbitration might be requested by either party to the 13(c) agreement. The term "labor dispute" is defined by the clause "to include, but not be limited to, any controversy concerning wages, salaries, hours, working conditions or benefits, including health and welfare, sick leave, insurance or pension and retirement provisions, the making or maintaining of collective bargaining agreements and the interpretion and application of such collective bargaining agreements, any grievance that may arise." The provision in a 13(c) agreement of arbitration for the settlement of all disputes is seemingly a case of the use of 13(c) agreements—intended by the legislation to protect labor from adverse effects arising from grant assistance—to affect the substantive terms of the collective bargaining relationship.

Extending Inefficient Work Rules under 13(c)

Work rules in the transit industry that have developed over a long period of history of collective bargaining have introduced inefficiencies in the deployment and use of labor resources. Section 13(c) prolongs and extends those inefficiencies into sequential periods when projects are undertaken with federal assistance. The section does so because it requires that a condition of the receipt of assistance shall be "the preservation of rights, privileges and benefits under existing collective bargaining agreements or otherwise," and "the protection of individual employees against a worsening of their position with respect to their employment."

The demand for transit service peaks twice daily in the early morning and late afternoon or early evening hours. It is reported that, typically, there are twice as many transit motor vehicles operating during the peaks than in the daytime hours between the peaks. The ratio of peak to off-peak operations is even higher than two for the hours of the day between the evening and morning rush hours.

In circumstances such as these, the efficient use of labor, which is a resource having valuable alternative uses for society, would suggest the widespread employment of part-time labor. In such an arrangement, many workers would be employed to perform transit industry service during peak periods and would then be released for the performance of services in other occupations and crafts during other hours of the day.

Such arrangements are not uncommon. Within the transit industry itself school bus drivers are employed for short periods of service each day and are then released to other activities. Similarly, casual labor is employed in many industries, workers are taken on for seasonal peaks and then released, and workers in many occupations are employed for the performance of defined daily tasks and are released upon their completion. When time-peaks and time-troughs occur in the demand for labor services, arrangements such as these imply efficient use of labor and the maximization of the economy's output, given the quantum of resources that society commands.

The transit industry would seem to be extraordinarily suited to the employment of part-time labor. In fact it employs very few part-time workers and it is inhibited from doing so by work rules that appear in the industry's collective agreements.[7] Those work rules sometimes put explicit numerical constraints, as well as a wide range of other restrictions, on the employment of part-time labor. These include minimum work hour guarantees for part-time drivers, maximum allowable times in a day or a week part-time drivers may actually be employed in vehicle operation, maximum number of hours per day for the interval between the beginning and the end of part-time drivers' service, and constraints on the place at which part-time drivers may begin and end their daily service.

These work rules forestall the employment of part-time personnel and encourage, instead, the employment of full-timers. Since part-time service is especially appropriate to an industry confronted with a widely fluctuating demand for its service, some transit labor is, in consequence, bid away from more highly valued service in other occupations and into transit employments in which part of their days are spent in relatively unproductive work or in idleness. The output of the economy is diminished, as is the productivity of the transit industry, and the operating costs of the transit industry are increased.

These work rules also have adverse effects upon transit industry efficiency. They increase the number of personnel that must be employed for a given level of service. They diminish the fraction of hours paid for that is actually devoted to vehicle operation. And they generate less efficient scheduling of service and less efficient deployment of labor.

Because it provides for the persistence of collective bargaining, per se, and of the rights, privileges, and benefits under existing collective bargaining agreements, section 13(c) tends to perpetuate and entrench work rules such as these, and, therefore, the inefficiencies they produce.

Constraints on Subcontracting

It is characteristic of many industries that some intermediate products and services and even some proportion of an industry's final product are not produced "in house" but, rather are purchased from supplying firms. This is sometimes called subcontracting. Firms that subcontract are, of course, motivated to do so because they can diminish their total production costs and achieve higher levels of overall efficiency by purchasing the services of supplier firms.

In general, the transit industry is foreclosed from subcontracting by 13(c) agreements. The "model" 13(c) agreement that has been used very heavily as a condition for the receipt of operating grants says that the recipient "shall be the sole provider of mass transportation services to the project and such services shall be provided exclusively by employees of the recipient." Services that had been previously secured from others through purchase, lease, or other arrangements are permitted to continue; but newly arranged subcontracting is not permitted.

---

[7] See Kenneth M. Chomitz and Charles A. Lave, *Part-Time Labor, Work Rules, and Transit Costs* (Washington, D.C.: U.S. Department of Transportation, 1981).

Thus new subcontracting that would diminish transit costs because subcontractors are more efficient producers of the relevant service is not permitted. Neither can transit systems avoid the payment of rents to their employees by subcontracting to firms with lower labor costs.

Labor Protection Standards

Section 13(c) does not prohibit the making of changes that will render the transit industry more efficient, even when those changes have adverse effects upon transit employees. Section 13(c) and the 13(c) agreements do, however, inhibit such changes by making them more costly than they would be in the absence of 13(c) protections. If the incremental costs are high enough, relative to the efficiency gains they will produce, such changes will not be carried out. That is to say, by increasing the cost of efficiency changes, 13(c) causes fewer changes to be made and prolongs some inefficiency.

Section 13(c) provides that arrangements be made to protect the interests of employees affected by assistance grants. Those agreements must include the protection of individual employees against the worsening of their positions with respect to their employment. As a result, section 13(c) agreements contain detailed, and sometimes elaborately defined, definitions of adversity effects and of the remedies that are intended to make whole those employees who are affected adversely by changes that are part of or incidentally produced by projects for which assistance has been given. Remedies include, among others, displacement allowances for workers placed in a worse position, dismissal allowances for workers laid off for as long as six years, the continuation of fringe benefits for dismissed employees during the period when dismissal allowances are paid, the payment of moving expenses in some circumstances, reimbursement for losses upon sale of a home or from the failure to occupy a leased dwelling for the full period of a lease, and lump-sum payments to employees who resign when they have received dismissal notice.

It can be seen that changes in the organization of transit services are made more costly by the labor protection constraints imposed by section 13(c). Principles of rational, maximizing management will, in these circumstances, foreclose some changes that would otherwise have occurred; less efficient methods will sometimes be preferred to more efficient methods because the latter will be encumbered by costs produced by the labor protection standards of the act.

Changes in the organization of work or in the level or organization of services performed by a transit system may diminish the number of employees or move some employees into a different occupational class or into a different period for the performance of service. It serves the interests of employees, therefore, if the area within which the labor protection provisions of the act are to be applied is given the broadest possible definition. Transit employees can be expected to make the claim that any adjustments that adversely affect them are project-connected, even if the connection is loose and amorphous.

Since the adverse effects tend to forestall change when labor protection arrangements apply, a broad interpretation of the area of effect of assistance projects can be expected to raise strong impediments to adjustments that affect the number of employees or the properties of their work assignments. Hence, a not uncommon prescription in 13(c) agreements is that "the term 'project' . . . shall not be limited to the particular facility assisted by federal funds, but shall include any changes, whether organizational, operational, technological, or otherwise, which are traceable to the assistance provided whether they are the subject of the grant contract, reasonably related thereto, or facilitated thereby."

The Labor Department's Role

It is possible that the Department of Labor has constructed administrative procedures that have intensified the employment of 13(c) as an instrument for the production of rents for the workers of the transit industry.

Section 13(c) says that "it shall be a condition of assistance . . . that fair and equitable arrangements are made, as determined by the Secretary of Labor, to protect the interests of employees affected by such assistance." The section then goes on to specify the minimum limits of the protective arrangements.

The Department of Labor has, in general, relied on the negotiation and signing of 13(c) agreements by the relevant transit managers and trade unions for the specification of the detailed protections that shall suffice for certification by the Secretary of Labor that the 13(c) conditions have been met. Only in rare instances does the secretary certify on his own initiative, and then only when the negotiators have not been able to reach agreement.

The Department of Labor apparently designed this procedure in response to its perception of congressional intent. The Senate Banking and Currency Committee Report on the Urban Mass Transportation Act of 1963 (88th Cong., 1st sess. Report no. 82, Calendar no. 69, p. 28) does say:

> The Committee does not believe that it is feasible to enumerate or set forth in great detail the provisions that may be necessary to assure the fair and equitable treatment of employees in each case. In this respect, it is expected that specific conditions normally will be the product of local bargaining and negotiation, subject to the basic standard of fair and equitable treatment. The Committee expects that the Secretary of Labor . . . will assume responsibility for developing criteria as to the types of provisions that may be considered as necessary to insure that worker interests are adequately protected against the kinds of adverse effects that may reasonably be anticipated in different types of situations.

The congressional intent to be inferred from this passage in the committee report seems capable of supporting reasonable alternative interpretations for the design of administrative practice. The inference can be drawn that the Congress intended local negotiation of the details of labor protection provisions as an a priori condition for the grant of assistance. But another, equally reasonable interpretation of congressional intent is that the Secretary of Labor certifies the general fulfillment of the labor protection conditions of the statute on the initiative of his own examination of the circumstances, and that detailed and specific labor protection provisions be the subject of local bargaining and negotiation, after the grant of assistance has been given. While it is clear that the Congress intended that specificities and details be locally negotiated, the Congress did not seem to specify at what point in the process the negotiation would occur.

By constructing the administrative process so that, generally, negotiation and agreement must take place prior to the secretary's certification, section 13(c) was transformed from an instrument for the mere protection of labor from adverse effects of transit assistance to an instrument for enlarging the power of the unions of transit employees to extract rents in the payment for their services. Since assistance to transit systems cannot be given until the unions have agreed to sign a 13(c) agreement, the unions are able to make their willingness to sign without protracted delays conditional upon understandings about the terms of collective bargaining agreements when they are next negotiable.

The process of transit grant assistance involves a time sequence of operations. The point in this sequence at which detailed local negotiations are placed can be expected to affect the consequences produced by the negotiations. Of the alternatives that are available, the prudent course for the Department of Labor would have been to arrange certification procedures so that they would tend to achieve *only* the fundamental congressional intention of the Urban Mass Transportation Act. With respect to labor standards, this means that labor be protected from the possible adverse effects of transit assistance. Instead, the procedures were organized in a way that permits the 13(c) agreement negotiation to have enlarged effects beyond that purpose and to intrude, as well, upon the definition of the terms of employment independently of the effects of transit assistance.

Mass Transit Viewed as a Public Service

The deficit of transit system operating expenses over operating revenue has been rising over time and the rate of increase of the operating deficit has been especially large in recent years. This has been a result of the more rapid rise of transportation and maintenance costs than of passenger revenue. The number of passenger trips has increased somewhat in recent years, but the mean fare charge to users has been quite stable and has, indeed, declined recently in real terms.

It is sometimes said that the rise in costs of operation and the consequential increase in the magnitude of the deficits, ought not be a matter of concern because mass transit is a "public service." The charge for its operation, it is said, is properly assessed upon the whole community, rather than users of the service. Thus, if it is appropriate that the cost of the system be borne from appropriations from the community's general revenue, it is a matter of indifference that deficits are large and rising. If the entire cost of a "public service" should be paid for by the public as a whole, it is surely proper that less than the total cost be covered by public subsidy. Therefore, rising deficits cannot be taken as a signal that costs are rising too quickly, nor do rising deficits suggest that steps should be taken to diminish operating costs.

This argument has no merit from an efficiency point of view because it involves a misunderstanding of the theory of public goods. It is true that price theory instructs that if an activity produces a public good, it should receive a public subsidy in order to secure an efficient allocation of resources. For such public goods, competitive markets will generate underinvestment, and subsidies are necessary to secure a socially optimal size for the industries that produce them.

A public good has the properties that it is impossible, or administratively very difficult and costly, to exclude anyone from consuming it and that its consumption by one does not diminish the quantity of it that is available for consumption by another. Mass transit services, however, are not a public good. The fare box is an efficient exclusionary instrument. Those who pay the fares are admitted and those who do not are excluded from the consumption of the service. When a unit of the service is consumed, it ceases to be available to others. The occupancy of a seat by one deprives another of its use; increments of transit service are not produced at a zero cost. It follows, since transit service is not a public good (even if transit systems are publicly owned), that user-fee revenue systems will generate a socially optimal scale and a socially optimal allocation of resources in the mass transit industry. This means that if public subsidies are paid to the mass transit industry, they need to be defended on income-and-wealth-transfer grounds and not on "public service" grounds; subsidies may, of course, produce scale suboptimalities.

It flows from this line of argument that transit operating costs are, indeed, an object upon which it is appropriate to concentrate attention. It is desirable that costs be diminished because the magnitude of operating costs affects the magnitude of the transfer payments that are made by subsidy systems from non-users of the transit system to users and from light users to heavy users of its services. Therefore, if section 13(c) increases the operating costs of the system, it is proper that its effects be examined.

Securing a Proper Accounting of 13(c) Costs

A number of studies of the effects of 13(c) have been undertaken.[8] Some of them have found that, while generalized complaints have been made by transit system managers, few have been able to specify in detail changes in systems operations that would have been undertaken, but that were forestalled by fear of 13(c) costs.

---

[8] See, for example, Frederic B. Siskind and Ernst W. Stromsdorfer, *The Economic Cost Impact of the Labor Protection Provisions of Section 13(c) of the Urban Mass Transportation Act of 1964* (Washington, D.C.: U.S. Department of Labor, May 1978); U.S., Comptroller General, *Program to Certify the Agreements to Protect Employees Affected by Grants Made under the Urban Mass Transportation Act of 1964* (Washington, D.C., 1976); Jefferson Associates, *Report to the U.S. Department of Labor: Administration of Section 13(c)—Urban Mass Transportation Act* (Washington, D.C.: Jefferson Associates, 1972); James L. Stern et al., *Labor Relations in Urban Transit* (Madison, Wis.: Industrial Relations Research Institute, 1977); Robert C. Lieb, *Labor in the Transit Industry* (Washington, D.C.: U.S. Department of Transportation, 1976).

It is not difficult to resolve this paradox. In recent years, almost the whole of capital cost and a very large fraction of operating cost of the mass transit system has been covered by public subsidies which have been growing over time. This has been coupled with political pressure for the expansion of service and the promotion of the idea that it is singularly important to diminish the consumption of fossil fuels and to improve environmental quality, and that mass transit services are efficient instruments for the achievement of those purposes.

As a result, both the management and the unions in the mass transit industry have found a coincident interest in the expansion of transit services. This is true even if it implies running conventional, scheduled services on lightly traveled routes in lightly traveled periods of the day, the cross-subsidization of some transit passengers by others, and the performance of service by conventional systems that might be more efficiently subcontracted or performed by others if conventional systems withdrew from some service sectors and left gaps to be filled by a transit service market. In addition, if the public authorities are perceived to be willing to increase the magnitudes of subsidies as operating deficits become larger, system managers, given their product market monopoly, will be less reluctant to consent to increases in worker compensation than if the industry were expected to cover its costs with its fare revenues or if it were operating in a competitive product market.

In sum, transit system management operates in a world which prods it with only weak incentives to make changes that would tend to achieve increments of efficiency and would tend to displace labor or otherwise to produce adverse effects for labor. Since the structure of incentives in the industry tends to forestall changes that would have adverse labor effects, the managers of the system are hard put to suggest changes they would have made but did not make because 13(c) costs would have been incurred.

To secure a proper accounting of 13(c) costs, it is inappropriate to ask system managers, motivated by weak efficiency incentives, whether 13(c) forestalled changes and imposed costs upon transit system operations. More appropriately, the accounting of 13(c) costs should be responsive to other questions. These include: What is the magnitude by which payments to mass transit employees exceed the levels of payments that would occur if 13(c) did not reinforce the collective bargaining mode? What changes in the forms of organization of labor services in the industry would occur if 13(c) did not reinforce the collective bargaining mode? What is the magnitude by which, given the current levels of service, the number of workers employed by the system exceeds the number that would be employed if a different set of work rules instructed organization of work? To what extent would part-time labor or labor of less skill replace full-time labor and labor with more skill if a different set of work rules than that enforced by the collective bargaining mode were to become effective? To what extent would conventional transit service shrink if efficiency incentives were stronger? And by how much would the magnitude of that shrinkage be diminished because 13(c) would impose costs upon the system if shrinkage occurred?

The cost to society of section 13(c) is the increment in the quantity of real resources, of which labor resources are the most important component, that is consumed by the transit industry and, therefore, rendered unavailable to other industries, as a result of the effects of section 13(c). They include additional resources consumed by higher levels of service that are induced by section 13(c), as when a lower-level service requiring a smaller number of employees or part-time employees may be introduced with grant assistance under the act but is made more costly by section 13(c) prescriptions. They include, also, the additional resources consumed at given levels of service as a result of less efficient arrangements than would otherwise have been made, had they not been forestalled by section 13(c).

The money cost to transit systems of section 13(c) is the increment in operating expenses under 13(c) constraints over what those expenses would have been in the absence of that section.

Both of these costs are extraordinarily difficult to quantify. This is so because, if there were no 13(c) prescriptions to fortify collective bargaining in the industry, assist in the production of rents in the industry's labor market, and make the adjustment of transit industry practice more difficult, departures from competitive market incentive structures would still survive. These include the survival of the collective representation of the workers of the industry, the exercise of monopoly power in the labor market, the existence of rents paid to labor, the diminished incentives of managers of publicly owned and subsidized properties administering a monopoly in the sale of scheduled transit service to reduce costs and introduce more efficient methods, and the political impediments that make it difficult to shrink service levels and raise user fares.

In these circumstances it can be expected that even in the absence of a section 13(c) the resources employed by the transit industry would be larger, the level of efficiency of its operations lower, and its money costs higher than if competition existed in both product and labor markets, and if the managers of the industry were required to secure revenues from users of its services that would cover costs and provide a normal return on investment in the industry's assets.

The difficulty in measuring the cost of section 13(c), thus, derives from the difficulty of estimating the scale of the industry, the forms of its organization, the levels of its service, and the payments that would be made by it for the inputs it purchased, in a state of the world in which the enumerated conditions diminish competition in markets and diminish the role of profit-maximization and cost-minimization principles in instructing management behavior and choice among alternatives.

Efficiency and Equity Aspects of Job Protection

A price-theoretic analytical treatment of labor protection provisions associated with changes in public policy raises serious questions about such protection on both efficiency and equity grounds.

Labor protection arrangements are a tax on change; they specify, in the urban mass transit assistance case, that assisted projects shall not have adverse effects on employees currently employed in transit enterprises or, if they do, then adversely affected current employees shall be reimbursed for the losses they suffer from changes that are project-induced. By imposing costs on change, such an arrangement diminishes the quantity of change that will occur, impedes the achievement of more efficient organizational forms, and misallocates resources employed in the production of transit services.

Principles of efficiency do not, however, completely instruct the design of public policy. The community may be prepared to experience some negative increment of efficiency that is generated by a public policy, if that policy tends toward the achievement of some equitable purpose. In a trade-off such as this, attention should be given, of course, to the magnitudes of the variables that are involved in the exchange. Even if the community seeks to promote equity as well as efficiency, a policy might be rejected if the efficiency losses are large and the equity gains are small.

But the equity argument for labor protection is questionable. Pragmatically, labor protection arrangements advantage incumbent employees of transit systems but disadvantage other workers. Such arrangements give priority rights to incumbents at the expense of other workers who aspire to transit employment directly or, indirectly, through subcontract agreements. Job protection provisions in the transit industry also give compensation rights to workers who are adversely affected by transit assistance, but not to workers who are adversely affected by other causes. For example, no job protection is afforded to workers who are employed in a declining industry or whose skills have been rendered obsolete and their capital values diminished by the progress of technology. Finally, labor protection arrangements aid transit workers who are adversely affected by assistance grants, but they provide no protection for owners of physical assets who might be adversely affected by those grants.

Firms are given the option of continuing to operate according to prevailing practice with respect to the organization of work or of introducing change, subject to the constraint that currently employed workers who are made redundant by change receive compensation payments. A tax is imposed on change. This tax produces disincentives that may make it not worthwhile to introduce change. The incidence of such a tax on change is distributed among consumers of the industry's product, owners of complementary resources employed in its production (including complementary labor), and substitute classes of labor that may have been employed in the now-estopped new process. It is also borne by the taxpaying community when the deficits arising from the compensation payments are covered from the community's general revenue funds.

Once it is seen that the gains of some are secured at a cost to others, it is not certain that equity is achieved by making incumbents the beneficiaries of policy. It is understandable, of course, that incumbents should combine to lay claim to benefits for themselves and, indeed, that they should assert multiple grounds in its defense. It is not surprising that each values his own interest more highly than that of others. But there is no reason why neutral persons should rank equitable priorities in the same way.

The equity argument for the protection of labor from the adverse effects of policy is also rendered questionable by the operation of labor market forces. If there are two classes of employments alike in all respects except that the prospective adverse effects from public policy are expected to be large for one of them and small for the other, a well-functioning market will arrange to pay a higher wage on current account for each hour worked to those in the employment that will suffer large adverse effects and a lower wage on current account to those in the employment that will suffer small adverse effects. Since the future is not known with certainty, the market makes estimates of future occurrences with uncertainty. This does not mean that the risks of adverse effects are not paid for in market transactions but only that some will be underpaid and others will be overpaid for those risks. There is no reason to expect systematic underpayment and, on average, workers will have been compensated for risk. The market will already have paid workers for the adverse experiences that they confront. Labor protection arrangements therefore involve compound payments for those adverse experiences. On this ground, too, the case for labor protection provisions is not a strong one.

In sum, if the protection of labor diminishes efficiency and this is not clearly offset by the service of equitable purpose, it is not certain that the protection of incumbent labor is consistent with the public interest.

RIGHTS, CONTRACT, AND UTILITY IN POLICY ESPOUSAL

Leland B. Yeager

Rival Criteria

This paper defends one version of utilitarianism against supposed alternatives to it as a policy criterion. First it reviews the natural-rights doctrine, contractarianism, and criticisms of each. Several sections state and defend a broad version of utilitarianism. Examples follow of writings about rights and contract that tacitly accept utilitarianism. A summary concludes the paper.

Natural Rights

Clear-cut examples of the antiutilitarian rights doctrine are rare. One reason, apparently, is that doctrines ostensibly of that kind are in fact tacitly utilitarian—a point developed later. Murray Rothbard (1973, pp. 23–25) observes that most libertarians have adopted natural rights and rejected both emotivism and utilitarianism as the foundation for their nonaggression principle. Their basic axiom is the "right to self-ownership," absolute property in one's own body. Accordingly, everyone has the right to perform actions necessary for surviving and flourishing without coercive molestation. Rothbard then develops a theory of property in nonhuman objects by appealing

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Ludwig von Mises Distinguished Professor of Economics at Auburn University. This paper combines material from three unpublished papers: "Utilitarianism and Rights in Policy Espousal," presented at a meeting of the American Association for the Philosophic Study of Society, Ann Arbor, May 1979; "Contract and Truth Judgment in Policy Espousal," presented in a seminar at George Mason University, October 1983; and "Utility, Rights, and Contract in Policy Espousal," presented at the meetings of the Midwest Economic Association, Chicago, April 1984. Yeager (1984/85) has briefly argued that the position expounded here is essentially, despite differences in terminology, the position of F. A. Hayek.

to John Locke's concept of a person's rightful ownership of previously unused natural resources that he first transforms by his labor (1973, pp. 26ff.). "These two axioms, the right of self-ownership and right to 'homestead,' establish the complete set of principles of the libertarian system" (p. 40).

Rights theorists reject the approach that would take a stand on each specific policy issue, such as deregulation of a particular industry or imposition of wage and price controls or government credit allocation, according to the apparent merits of the individual case; they reject narrowly focused cost-benefit calculations. Instead of being framed by case-by-case expediency, policy should conform to persons' rights.

And those rights should not be defended on a utilitarian basis, for doing so supposedly opens the door to all sorts of pragmatic demands for government intervention. The utilitarian, Rothbard complains (1973, p. 24), will rarely adopt a principle as an absolute and consistent yardstick. Instead, he regards it as a vague guideline or aspiration or tendency that may well be overridden. Milton Friedman, for example, although devoted to the free market as a general tendency, in practice allows many damaging exceptions to freedom against state intervention (Rothbard 1973, p. 24). Utilitarianism—or rather, in my view, its exaggerated pragmatist version—contrasts sharply with a pure doctrine of rights. Rights theorists derive positions even on quite specific policy issues from a very few propositions taken as axiomatic.

Unfortunately, the door is open to interventionist demands anyway. Libertarians cannot keep it closed by issuing methodological pronouncements or by reporting their intuitions about endangered rights. A pure rights position, untainted by utilitarian aspects, might serve for warding off illegitimate or undesirable interventions if it enjoyed general acceptance. Although it might be convenient if a particular doctrine were true and generally accepted, that convenience alone is no evidence, unfortunately, that the doctrine is in fact true. To make the best of reality, which often is inconvenient, we must face it as it is.

In reality, no doctrine will automatically protect us from bad interventions. Abuse of utilitarian arguments cannot be prevented by rejecting utilitarianism *tout court*. Instead, one must enter into discussion with one's pragmatist opponents, demonstrating how acting on an excessively narrow utilitarianism violates rights as intelligibly conceived and convincingly defended and so impedes the pursuit of happiness.

Rights and Value Judgments

What I am rejecting is a *pure* rights doctrine scornful of any utilitarian underpinning. I accept a pro-rights doctrine, provided that propositions about rights are recognized not as purely positive propositions of fact and logic but rather as normative propositions. A definite list of rights is unnecessary here. Suffice to say that rights are persons' entitlements to freedom from coercive interference by their fellows and by government; they concern life, liberty, property, and the pursuit of happiness. (When embodied in constitutions or statutes, of course, normative propositions like that take on an additional, legal, status.)

To identify assertions of rights as value judgments is emphatically not to disparage them as mere expressions of emotion or whim. It is instead the claim of imprescriptible ontological status for them that disparages them by exposing them to easy ridicule, as in Jeremy Bentham's remark (1973, p. 269) about "nonsense upon stilts." Although they are normative, assertions of rights can be supported by appeal to facts of human nature and other aspects of reality and to the findings of psychology, economics, and other disciplines.

As the philosopher Paul Edwards (1965) and the economists Sidney Alexander (1967, especially pp. 105–7, 114–15) and A. K. Sen (1970, especially pp. 62–64) have argued explicitly and as many other writers have recognized, fact and logic can be brought to bear in trying to clear up disagreement over all but fundamental value judgments. (Supporting considerations appear in Adler 1970.) It can be a constructive enterprise to try to clear up disagreement over specific or nonfundamental values. (Examples are the judgments that Jones should be sent to jail, that lying, cheating, and stealing are wrong, that private property is a desirable institution, and that specified rights should be recognized and respected.) It is antiintellectual simply to chalk disagreement up to irresoluble emotional differences. We can give and discuss *reasons* for value judgments.

Reasons for and against specific value judgments might include positive analysis showing why accepting some tends to promote and accepting others to subvert a society of a kind conducive to its members' successful pursuits of happiness. An important objective element, utilitarian in a broad sense, thus enters into the vindication of human rights. Only when we finally have nothing left to appeal to beyond some such fundamental value judgment as one favoring happiness and abhorring misery have we exhausted the relevance of facts and logic, investigation and discussion. But bona fide disagreements in the real world seldom if ever center on fundamental values, openly avowed. For practical purposes of policymaking, then, the fact-value distinction fades away.

Abstract Reasoning in Favor of Rights

The most immaculately antiutilitarian version of the rights doctrine boils down all too soon to simply promulgating rights and even to anathematizing disagreement as a sign of moral deficiency. Some antiutilitarian proponents of rights (Rothbard 1973, Rothbard 1982, Gewirth 1978, Mack 1978, Paul 1978) do offer arguments, after all, but arguments of peculiarly abstract kinds—appealing (as already noted) to self-ownership and Lockean homesteading or, alternatively, to what one must say about rights to avoid logical error.

Consider, for example, what Alan Gewirth does with his Principle of Generic Consistency, which, "unlike utilitarian and material deontological theories, . . . contains within itself the ground of its necessity; it is self-justifying" (1978, quotation from p. 203). Certain rights must be respected if each person, a purposive being, is to strive effectively for his purposes. He cannot consistently claim these rights for himself while denying them to persons affected by his actions when the very reasons he gives for claiming the rights for himself also apply to those other persons. He would be uttering logically contradictory propositions. In effect he would be saying: All persons for whom such-and-such reasons hold, including me, have such-and-such rights; yet he would deny, when expedient for himself, that other persons have those rights even though the stated reasons do hold for them also.

I question the supposed logical contradiction. An egoist might consider it expedient to claim certain rights for himself and deny them to others when he can get away with it. *He* is not necessarily accepting certain statements about rights as objectively true; *he* is not committing logical error. Instead, he considers it expedient to encourage other persons to hold certain beliefs about rights, beliefs in which those others might perceive internal contradictions if they were astute enough. If he can thereby further his own purposes, why should he care about the contradictions in other persons' beliefs? (In taking up what he calls "the Machiavellian case," pp. 196–98, Gewirth does try, but ineffectually, to rebut a counterargument similar to but not the same as mine.) We might consider the person a scoundrel, but that is not the same as his being a poor logician.

Eric Mack (1978) tries to derive the existence of rights from the proposition that coercion is deontically wrong—wrong because of its very character and not just because of its consequences. Because each person is an end in himself, it is deontically wrong for others to cause his actions to be out of accord with his own purposes, which is what coercion does. The deontic wrongness of coercion condemns any destruction of freedom by it. Contractual rights and property rights can also be vindicated by recognition of their violation as essentially coercive. Mack recognizes that utilitarian reasons for respecting rights could be grafted onto his argument but maintains that his argument does not depend on them.

Arguments, Not Just Intuition

The examples already given illustrate how antiutilitarian ethicists stop short, before a utilitarian would, with mere propounding or with appeal to irreducible intuition (even if they do not use the term). But what is intuition based on? Perhaps on an unarticulated recognition of consequences—tacit utilitarianism. Conceivably humans have an innate propensity, elaborated by natural selection, to develop such intuitions. If so, the consequences of having or not having them must have figured in the biological process.

Why not, then, strive to clarify those intuitions by disciplined observation and reasoning? It seems inadequate to rest everything on the assertion, for example, that individuals, as ends in themselves, simply should not be coerced. Such a line of argument does not take adequate account of the social context in which questions about rights arise. The utilitarian, however, does press on with an empirically oriented investigation into what sorts of institutions and practices do and what sorts do not accord well with human nature and the human condition.

Supposed axioms about rights cannot serve as the ultimate foundation of one's conception of desirable social arrangements. Instead, propositions about rights must be argued for, along with other propositions about what makes for a good society. John Gray (1983, especially pp. 59–60, 66, 68) maintains that John Stuart Mill in effect sought—successfully, Gray implies—to provide a utilitarian underpinning for respect for rights.

Another ethicist-economist who does so is Henry Hazlitt (1964, pp. 286–87). Within necessary qualifications, he says,

> legal rights are or ought to be *inviolable*. And so, of course, should moral rights be.
>
> This inviolability does not rest on some mystical yet self-evident "law of nature." It rests ultimately (though it will shock many to hear this) on utilitarian considerations. But it rests, not on *ad hoc* utilitism [sic], on expediency in any narrow sense, but on *rule* utilitism, on the recognition that the highest and only permanent

> utility comes from an unyielding adherence to *principle*. Only by the most scrupulous respect for each other's imprescriptible rights can we maximize social peace, order, and cooperation.

Elsewhere Hazlitt (1978, pp. 22–23) describes natural rights as "simply the rights that people ought to have." He notes the idea that rights gain in sanctity and respect by being called "natural," as if they were "something built into the universe, prior to creation, prior to existence." Actually, "natural rights" is "a mystical phrase. It's simply an unnecessary concept."

Before noting expositions of the rights doctrine that are tacitly utilitarian, we must consider the latter doctrine. First, though, a review of another rival doctrine is in order.

Contractarianism

A "contractarian" approach or attitude toward public policy has won respectful attention in recent years. Its most forceful and prolific advocate has been James M. Buchanan. (See the bibliography and, for some criticism, Gordon 1976 and Samuels 1976). Buchanan and other contractarians often also cite John Rawls (1971) with approval.

Contractarians exalt the individual over "society," agreement over coercion, and application of consent as the overriding criterion of desirability not merely to small-scale interpersonal relations but also in the large—to the broad framework of social, political, and economic institutions. A social contract—if not an actual one, at least a "conceptual" one—figures prominently in their vision.

Quotations from James Buchanan will help convey the contractarians' case and their objections to the allegedly opposed "truth-judgment" approach.

> To the contractarian that law is legitimate, and just, which might have emerged from a genuine social contract in which he might have participated. That law is illegitimate, and unjust, which finds no such contractual basis. [1977, p. 127]
>
> My point is mainly that of emphasizing the use of process, as opposed to end-state results. . . . For Rawls, as for contractarians generally, that which emerges from contractual agreement is just. . . . [1977, p. 61]
>
> That is "good" which "tends to emerge" from the free choices of the individuals who are involved. It is impossible for an external observer to lay down criteria for "goodness" independent of the *process* through which results or outcomes are attained. The evaluation is applied to the means of attaining outcomes, not to outcomes as such. [1975a, p. 6]

Many more passages of similar import can be found in Buchanan's writings. He would have us approve or disapprove of states of affairs or sets of rules not by considering their substance but overridingly by appraising the process employed to reach decisions about them.

Some Preliminary Doubts about Contractarianism

That idea seems odd to me. It resembles the idea that whatever people freely choose is in fact good for them. (Broome 1978 exposes this fallacy. One of his examples concerns "Jane," who, perhaps out of a sense of duty somehow absorbed from her surroundings, chooses to sacrifice an independent life of her own to care for her aged mother. It does not necessarily follow that Jane's free choice best serves her own interest or fulfillment or happiness or even that it maximizes her and her mother's utilities combined. Broome is simply warning against an invalid inference, of course, and not saying that some authority should impose on Jane the lifestyle it thinks best for her.)

Price theory and welfare economics may legitimately, for some purposes, assume that people make choices so as to maximize their utilities on the basis of definite utility functions. It does not in the least follow, however, that freedom of choice is the very criterion of what to choose. A libertarian might deplore forcible interference with the use of addictive drugs. Yet this attitude does not commit him to the view that drugs serve the happiness of those who choose to take them. He would not be inconsistent in considering the drugs harmful and their use deplorable.

Similarly, voluntary agreement is not itself the criterion of what to agree on. To value voluntary agreement and the democratic process in no way commits an economist or social philosopher to value whatever institutions or policies such processes may grind out. Why should he feel obliged to withhold any criticism? Decisionmaking procedure is itself properly an object of approval or disapproval, but it cannot sensibly be taken as the sole criterion of what to decide.

Rather than suppose that proper procedure exhausts the content of the good society, it would seem reasonable to emphasize proper procedure as an important part of that conception. But it is hard to see how a procedure can be deemed good utterly apart from some consideration of its likely results.

As Scott Gordon observed in reviewing *The Limits of Liberty*, Buchanan was trying to derive "moral principles without the aid of any moral premise" by carrying over positive analysis of collective decisionmaking into fundamental political philosophy. Buchanan might reply that he does have a moral premise, a weak one favoring individualism. Anyway, in Gordon's view (1976, p. 583), Buchanan's attempt to get normative conclusions from analysis that is entirely (or almost entirely) positive "cannot succeed. . . . [T]ry as one will, that troublesome word 'ought' cannot be excised from political philosophy and no degree of sophisticated 'is' can take its place."

Striving for clarity may justify some harshness: except in brief and untypical passages, Buchanan tries to conceal his employment of and to shirk his responsibility for values that he, like anyone, *must* be employing when he recommends anything, even when what he recommends is process rather than substance as the criterion for appraising institutions and policies.

Viktor Vanberg (1985) also raises apt questions about a supposedly purely procedural criterion. Can we appraise a process or a set of rules solely by the procedure of establishing and changing it, and so on? How do we avoid infinite regress? Isn't some substantive criterion needed somewhere? Furthermore, doesn't an "agreement test" unavoidably muddle together observations of people's preferences and of their theories about how alternative institutions would work? (Vanberg attributes the latter point to Karen Vaughn. As Vaughn has also said somewhere, or so I am told, choosing constitutional principles is not as easy as choosing a toaster.) One implication, I should think, is that economists and social philosophers have a legitimate role explaining and preaching what social arrangements they consider best, and why.

"Contractarianism gets its name," says one commentator (Pettit 1980, pp. 147–48), "from the device which it uses to filter out people's enlightened preferences." It identifies just social arrangements as those that answer to people's preferences—not the sort recorded by ordinary voting, however, but rather the preferences that people would have if they were not influenced by narrow self-interest and bias.

But on what grounds would people—people in real life and even or especially people stripped of distinctive self-interests (as by being placed behind a Rawlsian "veil of ignorance")—prefer one set of social arrangements over others? On what basis can a social philosopher hope to distinguish between laws or constitutions that could and those that could not have been agreed upon under either actual or idealized conditions? The only plausible basis, I submit, is that people stripped of narrow self-interest would consider the preferred arrangements as coming closer to an impartial conception of the good society, that is, as being more conducive to happiness, than the alternatives. A contractarian might say: No; impartial people would prefer those arrangements because they were *fair*. But why is fairness so desirable? Either because it just *is*, because a compelling intuition insists on it, because the question is not further discussible; or else because people's treating each other fairly is conducive to their effective pursuit of happiness.

I assert not that contractarianism is flatly wrong but that it disguises its affinities with utilitarianism by repulsive and unnecessary fictions (about which I shall have more to say).

The Attack on "Truth Judgment"

Yet Buchanan sharply contrasts his contractarian vision with the "truth-judgment" approach, as he disparagingly labels a position akin to the utilitarianism defended in this paper. This scorned approach tends to "assume that there is a unique explanation, a unique set of rules which defines the elements of a good society and which, once discovered, will come to be generally accepted by informed and intellectually honest men" (1977, p. 75; the phrases quoted occur in interrogative sentences, but the context leaves little doubt that Buchanan is characterizing the approach he condemns).

Buchanan distinguishes further between his own approach and its rival. His contractarian conception is analogous to market activity, a search for agreement to the mutual benefit of the participants. The truth-judgment approach is analogous to the deliberations of a jury (1984, pp. 29–30). Earlier (1975a, p. 164) he had distinguished, similarly, between the politician and the judge. The politician seeks consensus and acceptable compromises.

> He is not engaged in a search for some one "true" judgment, and he is not properly behaving if he seeks to further some well-defined ideal drawn from the brains of his academic mentors. The judge is in a distinctly different position. He does seek "truth," not compromise.

As applied to politics, Buchanan deplores the judge or jury conception. It conduces to intolerance by those who think they have attained political truth. Implicitly they claim the right to impose truth on persons mired in error (1977, pp. 76–77). Their view entails

> a demonstrated willingness to impose nonvoluntary changes on the existing pattern of entitlements in social order. . . . Once truth is found, there is no moral argument to be raised against its implementation. Consent is meaningless in this context. Opposition can be variously characterized as stemming from ignorance, folly, or the exercise of selfish interest. In any case, the views of those who actively oppose the truth-carrying zealots are not treated as worthy of respect. And any requirement to compromise with such views arises only because the reformists might otherwise lack the power to impose 'truth' unilaterally. [1977, pp. 143–44]

Buchanan (1975a, p. 167) sees many social philosophers exhibiting

> intellectual and moral arrogance. An attempt to describe the social good in detail seems to carry with it an implied willingness to impose this good, independently of observed or prospective agreement among persons. By contrast, [his] natural proclivity as an economist is to place ultimate value on process or procedure, and by implication to define as 'good' that which emerges from agreement among free men, independently of intrinsic evaluation of the outcome itself.

Passages abound in which Buchanan dwells on this theme of the arrogance, the itch to play God, of those who presume to employ their own value judgments in trying to frame a coherent conception of a good society.

For contractarians, process and consent, not outcome or substance, form the criterion of goodness or desirability in human institutions and relations. " 'Truth,' in the final analysis, is tested by agreement. And if men disagree, there is no 'truth' " (1977, p. 113). "A scientist may advance an argument to the effect that a proposition is 'true.' His argument . . . may succeed in establishing a consensus among his fellow scientists. But the 'truth' of the proposition emerges only from this agreement and not from some original objective reality" (1977, p. 145n.).

Authoritarianism, Relativism, Fallibilism

This extreme relativism is remarkable, yet Buchanan finds adopting it necessary to avoid authoritarianism with regard to truth. Actually, these are not the only alternatives.

Harry Davis (1967/68) has described a third position distinct from the two between which Buchanan evidently thinks the choice must lie. The first of those is the authoritarian position of one who believes he possesses an infallible pipeline to objective truth. That is the position of the baseball umpire who insists that he calls balls and strikes as they objectively *are*. The second is a relativist-nominalist position of rejecting all absolutes and embracing a radical relativism or skepticism. An umpire holding this view says that pitches are neither balls nor strikes until he calls them. The third position, "fallibilism," combines metaphysical or ethical objectivism with epistemological relativism. The fallibilist umpire says, "I call 'em as I see 'em." On this view, it makes sense to seek objective truths about reality, even including truths about what is morally valuable and politically desirable. Yet no person or group is entitled to claim infallible possession of such knowledge. Each searcher contributes what he can, aware that his contribution is incomplete and perhaps erroneous. In science, culture, and philosophy, fallibilism calls for free discussion open to the competition of all ideas, evidences, and arguments. (Davis recognizes fallibilism as a central concept in the philosophy of Charles Sanders Peirce. See also Wiener 1968 and Peirce 1955.)

Far from being subversive of constructive discussion, a willingness to state clearly what one believes and why, exposing one's views to inspection and possible refutation, is essential to it.

The fallibilist position adopts the scientific attitude and method. Belief in the meaningfulness or at least the heuristic value of the concept of objective truth to be sought through research and discussion need not entail arrogance, elitism, and an eagerness forcibly to impose one's beliefs. Belief, for example, that one type of society is more conducive to human happiness than another in no way entails an eagerness to implement one's vision with force. A concern for process and for how decisions are made and implemented, an extreme aversion to having policies, even good ones, rammed down one's own throat or down other people's throats, may well be a major element in one's conception of the good society. An adherent of the truth-judgment approach may well harbor this strong concern for due process; it is not the private property of the contractarians.

Yet the contractarians tend to suppose that a policy not commanding a strong consensus in its favor (or whose adoption would not at least be in accordance with an underlying constitution deemed to command unanimous consent) is thereby revealed to be bad or undesirable. (Passages to this effect in Buchanan and Tullock 1962 are quoted in Yeager 1978, p. 200 and note 17.) The contractarians overlook a vital distinction: What is undesirable in such a case is not necessarily the policy itself but rather its imposition by antidemocratic means. It is not necessarily inconsistent or antidemocratic or elitist for an economist or social philosopher, while deploring imposition of the policy, to continue advocating it and trying to explain its virtues. In short, the nihilistic relativism of the contractarians is by no means the only alternative to a repulsive authoritarianism.

Contractarian Fictions

Contractarians understand, of course, that specific policy measures cannot in practice be enacted only with unanimous consent. They therefore try to find constitutional authority for nonunanimous enactments. The constitution need not necessarily be a written document; it may consist of the fundamental features of the existing society. That broad constitution is deemed to command substantially unanimous consent by virtue of its actual existence and people's living under it. (See, for example, Buchanan and Tullock 1962, pp. 254, 260–61.) Yet David Hume had long before mocked such notions of passive, tacit consent with his much-quoted example of the shanghaied sailor who, merely by refraining from jumping overboard, does not thereby consent to the captain's supposed authority over him ("Of the Original Contract," reprinted in Hume 1965, p. 263).

Buchanan and Tullock (1962, pp. 260–61) conceived of the "social contract" as

> a dynamic one. . . . We do not conceive the 'constitution' as having been established once and for all. We conceive the contractual aspects to be continuous, and the existence of a set of organizational rules is assumed to embody consensus. We think of the individual as engaging continuously both in everyday operational decisions within the confines of established organizational rules and in choices concerned with changes in the rules themselves, that is, constitutional choices. The implicit rule for securing the adoption of changes in these organizational rules (changes in the structure of the social contract) must be that of unanimity. This is because only through the securing of unanimity can any change be judged desirable on the acceptance of the individualistic ethic.

In later writings Buchanan also treats a supposed implicit social contract as if it had actual force. He refers (1975a, p. 96) to the "existing and ongoing implicit social contract, embodied and described in the institutions of the status quo. . . ." At greater length (1975a, pp. 84–85), he argues that the status quo must be evaluated as if it were legitimate contractually, even

> when an original contract may never have been made, when current members of the community sense no moral or ethical obligation to adhere to the terms that are defined in the status quo, and . . . when such a contract, if it ever existed, may have been violated many times over. . . . Does the presence of any one or all of these negations remove legitimacy from the status quo?
>
> Again it is necessary to repeat the obvious. The status quo defines that which exists. Hence, regardless of its history, it must be evaluated as if it were legitimate contractually. Things "might have been" different in history, but things are now as they are.

The interpretation conveyed by the foregoing quotations finds support in a book (1983) by one of Buchanan's former students. The social-contract theory of the state, Randall Holcombe explains, is an attempt to describe the legitimacy of the government's power. It views society as "a type of club, where all individuals conceptually agree to become members and adhere to the club rules." Actually, individuals are born into society and must adhere to its rules whether they agree to or not. "Here, the social contract theory of the state must fall back upon the conceptual agreement of all members of society. The state operates as if all members of society had agreed to its rules—as if there is unanimous approval of the constitution." (Holcombe 1983, pp. 124–25; compare passages of similar import on pp. 9, 123, 125–26, 134.)

Holcombe does not accept Buchanan's formulations wholly without reservation. "Since all of the members of the society did not actually agree to a social contract, . . . some type of conceptual agreement must be fabricated if the theory is to have any connection with reality" (1983, p. 155). The words "must be fabricated" deserve emphasis.

The writings of Buchanan and other contractarians (including Holcombe 1983, especially chapter 8) bristle with words like "conceptual" and "conceptually"—"conceptually agree," "conceptual agreement," "conceptual social contract," "conceptual unanimous approval," and the like. The very use of the words indicates that a "conceptual" agreement is not an actual one, that a "conceptually" true proposition is not actually true. It is no mere joke to say that "conceptually" is an adverb stuck into contractarians' sentences to immunize them from challenge on the grounds of their not being true.

Buchanan (1975b, pp. 123, 125) distinguishes between constitutional and postconstitutional stages of decisionmaking and envisages agreement or "conceptual" agreement at the constitutional stage as authorizing "apparent coercion" and "apparent redistribution" at the postconstitutional stage. By such fictions, realities like actual coercion and actual redistribution are interpreted away. They vanish into the realm of the merely "apparent" by being deemed in accordance with some agreement that may itself be merely "conceptual."

Even punishment, in Buchanan's view (1975a, p. 192n.) implements a contract. "In a genuine contractarian theory, there is no problem raised concerning the 'right' of some persons to punish others, since, in effect, individuals who find themselves in the implicit social contract that any social order presupposes have presumably chosen to be punished as the law directs when they violate law." Here the word "presumably" seems to mean "conceptually" in Buchanan's sense.

Further evidence of the contractarians' reliance on fictions is that Buchanan and many other commentators accept John Rawls's (1971) characterization of his own method as contractarian. Actually, no social contract at all is involved in Rawls's derivation of his principles of justice. Instead, Rawls employs elaborate fictions (about deliberations behind a "veil of ignorance") in more or less disguising—

perhaps even from himself—his total reliance on his own intuitions. (It astonishes me how many eminent scholars swallow Rawls's own characterization of his approach. Among the apparent minority who do identify what Rawls actually does are Hare n.d. and Gray 1978.)

No one need object to fictions if they are heuristically useful—if they stimulate the flow of ideas. Nor is it necessarily objectionable to employ fictions and figures of speech for expository and stylistic purposes. But a doctrine should not *depend* on them. Ideas that defy expression in straightforward, nonmetaphorical language incur deserved doubt by that very fact.

Contractarians might strengthen their case by occasionally presenting it, if they can, without resort to their favorite fictions. In doing so, however, they would be bound to erase sharp distinctions between their approach and the supposedly despicable truth-judgment approach. The version of the latter advocated in this paper does lead to much the same individualistic values as contractarianism, but its conceptual apparatus and expository style are quite different—more straightforward, and charier of fictions.

Utilitarianism Under Attack

Utilitarianism is routinely caricatured and scorned nowadays, and some versions do deserve scorn. It is said to be lowbrow, crass, and subversive of personal rights. The Benthamites, says Joseph Schumpeter (1954, pp. 133, 407–8), created "the shallowest of all conceivable philosophies of life."

That assessment might well apply if utilitarianism really did recommend that people spend their lives pursuing immediate pleasure. The utilitarianism defended in this paper, however, concerns the appraisal of social arrangements—policy espousal, as Philbrook (1953) would say.

Jeffrie G. Murphy (1977, p. 232) criticized a version that he attributed, probably wrongly, to John Stuart Mill:

> This theory is so obviously morally bankrupt that very few contemporary moral philosophers take it at all seriously. . . . [It] fails to pay attention to . . . important *autonomy* values . . . and thus fails to articulate a satisfactory conception of *justice* or *respect for persons*. It does not . . . rule out the sacrifice of persons for the general good.

Walter Grinder (1978, pp. 9–10) bewails "the tired and woefully pragmatic doctrines of end-state utilitarianism—the cursed Benthamism in all its permutations, that has proved the bane of liberty's existence for almost 200 years. During the 19th century, utilitarianism almost singlehandedly short-circuited the great classical liberal revolution." Frank S. Meyer (1962, pp. 1–2, 32–33) perceived a fatal flaw in the philosophical underpinnings of utilitarianism. Nineteenth-century liberalism "deserted its heritage of defense of freedom of the person" and "denied the validity of moral ends firmly based on the constitution of being. Thereby, with this denial of an ultimate sanction for the inviolability of the person, liberalism destroyed the very foundations of its defense of the person as primary in political and social matters." To utilitarians, "Human beings considered as the objects of operations are no more nor less than . . . objects. Kant's imperative is reversed. Our humanitarians of the welfare society take as their maxim: treat no person as an end, but only as a means to arrive at a general good."

The criterion of the greatest sum of the utilities of individuals is collectivistic, according to John Rawls; it regards individuals as processing stations for converting goods and services and experiences into increments to aggregate social utility. The ideal utilitarian legislator, in allocating rights and duties and scarce means of satisfaction, makes decisions similar to those of a maximizing entrepreneur or consumer; his correct decision is essentially a matter of efficient administration. "This view of social cooperation is the consequence of extending to society the principle of choice for one man, and then, to make this extension work, conflating all persons into one through the imaginative acts of the impartial sympathetic spectator. Utilitarianism does not take seriously the distinction between persons" (Rawls 1971, p. 27; cf. pp. 178–92, 449, 450, 572–73).

This last objection ties up with reference to the nonoperationality of the aggregate-utility criterion. Can one conceive of operationally meaningful rules for sacrificing individuals for the greater good of "society"? How good, how conducive to the pursuit of happiness, would a society with such rules be? And how would the utilities of different persons be measured and compared and added anyway? (Perhaps the clearest example of accepting the maximum-aggregate-utility criterion is Edgeworth [1881]. Edgeworth already pointed out [p. 136], as Rawls did later, that this version of utilitarianism requires extreme altruism; it connotes *"Vivre pour autrui."*)

Act-utilitarianism, as distinguished from rules-utilitarianism, is particularly objectionable. It calls on the individual to choose, in each separate case, the action appearing likely to contribute to the greatest excess of pleasure or happiness or good over the opposite. No notion of rights or principles should bar such a calculation, for respecting them is not an independent objective. Respecting them is fine when it happens in the individual case to serve the greatest excess of pleasure over pain, but that excess alone remains the final criterion.

Perhaps the clearest recent example of wanting each case handled on its own merits with no presumption in favor of respecting rights or principles occurs in Joseph Fletcher's *Situation Ethics* (1966). Fletcher departs from act-utilitarianism as ordinarily conceived only in making "love" rather than happiness the criterion and in making the altruism it calls for more blatant and cloying. Even this substitution makes little difference, since Fletcher interprets "love" as conduciveness to well-being, especially of persons other than oneself. In the coalition that he recommends between the love ethic and the utilitarianism he attributes to Bentham and Mill, "the hedonistic calculus becomes the agapeic calculus, the greatest amount of neighbor welfare for the largest number of neighbors possible." Fletcher "holds flatly that there is only one principle, love, without any prefabricated recipes for what it means in practice, and that *all other* so-called principles or maxims are relative to particular, concrete situations! If it has any rules, they are only rules of thumb." "The situationist holds that whatever is the most loving thing in the situation is the right and good thing" (quotations from pp. 95, 36, 65 respectively).

No wonder critics reject utilitarianism understood as something like that. It betrays remarkable arrogance to take it for granted that the actor in the individual case, tacitly endowed with the omniscience of the act-utilitarian or act-agapeic philosopher himself, can foresee all the immediate and remote and all the direct and indirect consequences of his actions, even including those working through reinforcing or undermining principles and habits and through affecting persons' moral characters, and can assess the good and bad values of all these consequences and strike a balance. As Peter S. Prescott (1973) has said, "what was recently called 'situation ethics' . . . can be defined as action based on invincible trust in one's own moral perspective."

An extreme act-utilitarianism or situation ethics might indeed countenance framing and executing an innocent man to pacify an angry mob and so avoid worse outrages—to mention the example so routinely trotted out (for example, McCloskey 1969, p. 181; compare Rothbard 1973, pp. 24–25, on the execution of all redheads to delight the rest of the population).

Some Answers to the Critics

John Rawls comments perceptively on this sort of attack on a strawman version of utilitarianism. (Rawls 1955, writing before McCloskey

1969 and before his own book of 1971, attributed the standard horrible example to E. F. Carritt. Rawls 1955 [pp. 76–78] also answered the question, raised in a similar vein, whether it is acceptable to break one's promise when the consequences of doing so appear good on balance. The very point of promising "is to abdicate one's title to act in accordance with utilitarian and prudential considerations in order that the future may be tied down and plans coordinated in advance. . . . The promisor is bound because he promised: weighing the case on its merits is not open to him." The institution of promising and promise-keeping itself has obvious utilitarian advantages. That the promisor's obligation may be overridden in exceptional hard cases does not mean that the obligation does not exist at all.)

How conducive to happiness would a society be, Rawls (1955) asks in effect, in which truth and rights were treated as contemptuously as in the hackneyed horrible example? More specifically, in what context might a judgment be made in favor of sacrificing the innocent victim to pacify the mob? It is vague to say that it would be a good idea if somebody did something to save many innocent lives by sacrificing one. How could such acts be institutionalized? Just who would be authorized, and in what circumstances and under what rules, to inflict "telishment" (by which Rawls apparently means ostensible punishment inflicted for ulterior purposes)? Several reasons are obvious, and Rawls suggests some, why an institution of telishment could hardly be justified on utilitarian grounds.

Does this dismissal of the telishment case simply postulate out of existence the difficulties that it is meant to illustrate? Am I denying that any single case could ever arise in which sacrificing an innocent person might appear to promise greater good on the whole? Well, can one ever be confident that such a case has in fact arisen? The consequences of supposing so and acting accordingly are unfathomable. Approving the violation of rights whenever the decision-maker thinks it would be beneficial on the whole would reinforce unhealthy temptations and undercut the very concept of rights. A society tolerating such violations would hardly be one in which people enjoyed relatively favorable opportunities to make good lives for themselves. Taking account of the associated institutions, habits, attitudes, and personality traits, as well as the fact that each person has a life and consciousness and purposes of his own, requires rejecting such a society. Endorsement of personal rights instead follows precisely on utilitarian grounds.

In the abstract, though hardly in convincing detail, one can contrive a case in which an act ordinarily deemed wrongful would have a net balance of good consequences, even with any undermining of respect for rules and rights counted on the negative side in the assessment. Perhaps the wrongful act can be kept secret, and its victim would have died soon of agonizing disease anyway. The contrived assumptions would rule out such adverse consequences as impairment of the agent's attitudes and moral character. The assumptions would render the proposition about the acceptability of the otherwise wrong act an empty tautology. *If* the act really would lead to the greatest net utility, absolutely all things accurately considered, and *if* the greatest net utility is one's criterion of what ought to be done, then the act ought to be done. But in what actual context could these *if*s be met? When could one have absolutely all the relevant knowledge of consequences, including the consequences of violation of valued principles, certainty that one's knowledge was accurate and complete, and certainty of one's accuracy in weighing opposing considerations? The assumptions required negate the real-world context that in fact recommends rules-utilitarianism over act-utilitarianism.

Suppose for the sake of argument, nevertheless, and quite implausibly, that a clear case of the postulated kind did arise. Suppose, further, that arranging compensation acceptable to the prospective victim or victims was for some reason impossible and that the burden of decision fell on me. Then I would have to face up to that case. I cannot commit myself in advance, ignorant of the specific facts, to insisting on observance of rules and rights even though the heavens should crumble. And neither could a self-styled antiutilitarian champion of rights.

A Defensible Utilitarianism

A sounder version of utilitarianism than the one routinely pilloried has been called rules-utilitarianism or indirect utilitarianism. (John Gray 1983, especially pp. 11–15, 31–32, 38–39, 46–47, attributes an indirect utilitarianism to John Stuart Mill and tries to distinguish it from rules-utilitarianism; for present purposes the distinction is inessential.)

According to McCloskey, however, rules-utilitarianism arises from awareness that the act version will not do; yet it is only pseudo-utilitarianism. It opts for "irrational" conformity to rules even though no intrinsic moral significance attaches to them and even when conforming to them brings greater total evil. The rules-utilitarian prefers conformity to a rule to maximization of good. If he replies that "his is the best way of promoting the greatest good, he is abandoning rule for act utilitarianism" (McCloskey 1969, p. 188).

Why, though, does McCloskey speak of *irrational* conformity to rules? Rules serve human welfare. Utilitarianism of course recognizes dilemma cases—they figure in the human predicament—in which applicable rules clash and in which some must be overridden to permit conforming to others more demanding in the particular case. To recognize such cases is not to lapse back into act-utilitarianism or situation ethics. The rules version does stress the advantages of habituation to rules and does caution against excessive readiness to override a rule (especially against the temptation to make an exception in one's own favor). If the idea of framing the innocent man to pacify the mob so shocks us, it is because we rightly find it so hard to imagine cases in which the rule of justice should be the one to be overridden. The quality of McCloskey's debating tactics is evident in his example (p. 191) of choosing between abiding by Australia's drive-on-the-left rule and colliding with a car in the wrong lane.

As for rules having no "intrinsic" moral significance, well, what is so suspect about there not being irreducible ultimates for which no arguments can be offered? Of course utilitarianism does not and cannot insist that rules be followed for their own sakes. On the contrary, rules are *instrumental* to a good society and thereby to people's happiness. To treat rules as absolute ultimates would undermine respect for them by making them appear ridiculous.

Let us consider the question of justice and injustice a bit further. Victimizing an innocent person or minority for the supposed greater gain of others is indeed unjust. But we need not stop short with reporting our intuition to that effect; we can give reasons for our judgment. We do not downgrade justice by not regarding it as an undiscussable ultimate. Injustice subverts social cooperation and the pursuit of happiness. Sir James MacIntosh argued (*Vindiciae Gallicae,* quoted in Halévy 1955, p. 185) that the extreme usefulness of general principles of justice makes them morally obligatory.

> Justice is expediency, but it is expediency, speaking by general maxims, into which reason has concentrated the experience of mankind. . . . When I assert that a man has a right to life, liberty, &c. I only mean to annunciate a *moral maxim* founded on *general interest,* which prohibits any attack on these possessions. . . . [A Declaration of Rights is an expedient] to keep alive the Public vigilance against the usurpation of partial interests, by perpetually presenting the general right and the general interest to the Public eye.

John Stuart Mill (*Utilitarianism,* Chapter 5, p. 299 in Mill 1968) said: "Justice is a name for certain classes of moral rules which concern the essentials of human well-being more nearly, and are therefore of more absolute obligation, than any other rules for the guidance of life. . . .”

Social Cooperation and Comparative Institutions

As an approach to understanding what sorts of individual conduct and especially what institutions and policies are desirable, rules- or indirect utilitarianism amounts to much the same thing as the comparative-institutions or good-society approach. These all have an affinity with the “truth-judgment” approach disparaged by James Buchanan.

For present purposes it matters little whether these different labels apply to exactly the same doctrine; if not identical, the doctrines do share a common orientation. Their adherents try to form a conception of the good society by contemplating and comparing alternative sets of mutually compatible social institutions. Their ideal is whatever arrangements best facilitate the success of individuals seeking to make good lives for themselves in their own diverse ways. (Strictly speaking, their ultimate criterion is human happiness, however best served. It is a researchable and discussable empirical judgment that happiness is served by institutions that facilitate voluntary cooperation, including ones that secure the rights mentioned in the U.S. Declaration of Independence.)

This approach recognizes the importance of mutually beneficial cooperation among individuals—through peace and security and through the gains from specialization and exchange. (Adam Smith pointed out that man is a social animal: he makes contracts. As Scott Gordon [1976, p. 586] notes perceptively, “the most important feature of that word is the final letter, which makes it plural. There is a world of difference between the conception of society as consisting of contracts and the conception of it as based upon a contract.”)

The approach recommended here appraises particular principles, rules, institutions, and policies according to whether they are likely to serve or subvert social cooperation in the sense just indicated. (The concept if not the term is prominent in the philosophies of Thomas Hobbes and, as F. A. Hayek has emphasized, of David Hume. See Hobbes, *Leviathan,* chap. 15, and Kemp 1970. The term “social cooperation” is prominent in the writings of Herbert Spencer, Ludwig von Mises, and Henry Hazlitt.)

Social cooperation counts as a near-ultimate criterion, since it is an indispensable means to individuals’ effective pursuit of their own happiness in their own diverse ways (Hazlitt 1964, especially p. 36). Cooperation is facilitated by rules that improve people’s chances of predicting each others' behavior and achieving coordination. Voluntary cooperation accords better than coercion with each person's having purposes and ideals of his own and with his having only one life to live. Emphasis on voluntary cooperation warns against authorizing any agency to impose unfair sacrifices on individuals for the supposed greater good of a greater number. But this approach does not simply postulate voluntary cooperation and deplore coercion. It investigates and compares the types of society likely to emerge from having alternative sets of institutions and rules and, in particular, from whether or not personal rights are recognized and respected.

The term "utilitarianism" as used here applies to any critical examination of social rules and institutions, their functioning, and their implications for happiness. In this wide sense, F. A. Hayek says (1976, pp. 17–18), anyone prepared to examine existing values rather than accept them unquestioningly is a utilitarian; Aristotle, Thomas Aquinas, and David Hume would so count.

In rejecting act-utilitarianism for rules-utilitarianism—but the terms are not his—Hayek explains (1969, pp. 45–46) why it may be rational to disregard known particular circumstances when making decisions. Accidental and partial bits of information might not change the probability that if we knew and could process all information about the circumstances, the net advantage would lie on the side of following the applicable rule. We should not decide each case on the basis of the limited number of individual facts that we happen to know.

One reason for abiding by rules, then, is that we simply cannot assess all the consequences and costs and benefits—direct and remote, immediate and delayed—of alternative actions in each particular case. One might object that this position is antiintellectual, making a virtue of ignorance. How can we know that advances in theory and technology may not make possible those allegedly impossible assessments? Part of the answer, I conjecture, is that the critic has not really seen the point. Rules-utilitarianism does not glorify ignorance. Rather, it perceives the rationality of acting, in certain cases and aspects of life, on generally applicable abstract principles instead of on the fragmentary and probably accidentally biased bits of concrete information that one may happen to possess. Furthermore, complexity and ignorance by no means form the entire case for rules-utilitarianism. Acting by rule or on principle often contributes to *overcoming* ignorance, namely people's ignorance of each others' probable behavior. General acceptance of principles contributes to predictability in the world and thus to people's chances of coordinating their activities to their mutual benefit.

As a utilitarian in the tradition of Hume rather than Bentham, Hayek does not envision maximization of some aggregate of numerical measures. He says (1976, pp. 129—30) that the aim in developing or altering rules of just conduct "should be to improve as much as possible the chances of anyone selected at random." He speaks of *chances* rather than *probabilities* "because the latter term suggests numerical magnitudes which will not be known." Equivalently, "the best society would be that in which we would prefer to place our children if we knew that their position in it would be determined by lot" (1976, p. 132; similar passages occur in 1967, p. 163; 1978, pp. 62–63; 1976, p. 114).

Hayek's formulations are similar to those of John C. Harsanyi, an avowed utilitarian, who considers a person contemplating alternative social arrangements in ignorance or at least in disregard of what his personal situation would be. On Harsanyi's theory, that person "would have to choose the social situation yielding him a higher expected utility, which in this case would mean choosing the situation providing a higher *average utility level* to the individual members of the society." (Harsanyi 1976, chap. 5, p. 67; cf. Harsanyi 1973, pp. 276–77. If Harsanyi's method resembles Rawls's [1971] notion of choice behind a veil of ignorance, the similarity goes to show that such a conception of impartiality need not be a distinctively contractarian one, as Rawls seems to think.)

The version of utilitarianism here attributed to Hayek, among others, might also, as already suggested, be called a comparative-institutions or good-society approach. Although Hayek repeatedly emphasizes how spontaneously evolved rules and institutions may serve an order that tends to reduce conflicts and ease cooperation among persons pursuing their own diverse ends, he does not discourage looking critically at those rules and institutions and sometimes deliberately modifying them.

Consciously designing a society from scratch, however, is not a live option. No one knows enough for such an undertaking. Through trial and error and survival of what works, our existing society incorporates much unarticulated knowledge. Throwing that knowledge away merely because of its being unarticulated and therefore unappreciated would be reckless. We should have a certain humility in undertaking reform—so Hayek in effect argues—but not reject all thought of reform.

Utilitarianism and Political Obligation

The comparative-institutions strand of utilitarianism does not try to ground government and political obligation in contractarian fictions. The notion of consent—tacit consent—may have heuristic value, admittedly; but arguments using it should, if sound, be amenable to translation into straightforward English. In such arguments, tacit consent alludes to considerateness, reciprocity, and self-esteem, all of which are valuable on broadly utilitarian grounds. Most of us believe that we should ordinarily practice the everyday little courtesies toward one another, accepting trivial inconveniences for ourselves to spare others substantial ones. Ordinarily, for example, we do not complain about reasonable noise incidental to useful activity ("reasonable" is admittedly a weasel word here); for we ourselves benefit from a society in which such noise is tolerated. The point is not that we and others have agreed, or are "deemed" to have agreed, to suffer noise. Such a fiction is unnecessary. Rather, each of us refrains from complaining about reasonable noise in the expectation of others' similar forbearance and in the interest of the social cooperation from which we all benefit.

Each of us would be uneasy constantly and belligerently insisting on our "rights," refusing consideration of others at the cost of even the slightest momentary inconvenience, and insisting that others either refrain from activities exerting the slightest adverse externality or else pay compensation. That would be a nerve-wracking way to live. We would be inconsistent in being intolerant and inconsiderate of others while expecting them to be tolerant and considerate of us. Such behavior would emit messages to others about our own character.

For most of us, furthermore, it serves our self-esteem to think of ourselves as consistent, considerate persons who play fair and who support rather than subvert a decent society. Each of us benefits from courtesy and ease in relations with our fellows. It does not serve our self-esteem or interest to undercut that spirit. Thus, we need not interpret tolerant and considerate behavior toward others as compliance with a contract.

Much the same considerations argue for respecting the legitimacy of government and an obligation to obey its laws. The argument also explains why the legitimacy and obligation are not total. Most of us feel obliged to obey a reasonably decent government on the grounds that doing so contributes to our own and our fellows' welfare—in view of the Hobbesian alternative. Consideration of our fellows, which ordinarily serves our self-esteem, requires our not contributing to the subversion of a generally useful institution (which government is, even though a "necessary evil"). Unfairly arrogating special privilege to oneself, picking and choosing which laws to obey, and making exceptions in one's own favor does something toward undermining the legitimacy and authority of a government from which we ourselves derive net advantages. (I distinguish, of course, between a government that is decent on the whole and, on the other hand, a tyrannical one that ought, on the same grounds of concern for ourselves and our fellows, to be overthrown.)

On grounds involving one's own and other persons' happiness as served through social cooperation, then, one can make a case for people's (1) according everyday courtesies to one another, and (2) practicing considerateness and reciprocity in yet another way, through respecting political obligation toward a reasonably decent government.

Much of the following ties in with possible answers to what Brand Blanshard calls the fundamental question of political theory: "Why should I obey the law? An adequate answer to that question would carry with it the answer to such questions as, Why should there be a government at all? What are the grounds of its rights against me and my rights against it?, and How in principle are those rights to be limited?" (Blanshard, 1961, chap. 14, "Reason and Politics," quotation from p. 376.)

According to the first theory that Blanshard reviews, political rights and duties are based on nothing. This anarchist view is "doctrinaire idealism of a pathetically irresponsible sort" (1961, p. 378). Second is the doctrine that might makes right; it sets ethics aside. The third appeals to divine authority, the fourth, the doctrine of Hobbes and Rousseau, to a social contract.

Contractarianism begs the question, Blanshard explains. If, before entering into the social contract, I do not have an obligation to keep contracts, then the social contract to keep future contracts is not binding, nor are future contracts supposedly made under it. But if I do have an obligation in the first place to keep contracts or to honor certain other duties, then the theory is superfluous.

Fifth is the doctrine of the Declaration of Independence: the self-evident truths that men have certain unalienable rights, that governments are instituted to secure these rights, and that if a government becomes destructive of them, the people have the right to alter or abolish it. This doctrine, says Blanshard, comes close to the correct one. He does not deny the existence of natural rights resting on no government and no convention and identifiable by reason, but he doubts that the doctrine of self-evidence states their true ground. Natural rights can break down: cases are conceivable in which the community can legitimately exercise coercion. All sorts of rights would be desirable—here I am embroidering on Blanshard a bit—if recognizing them did not cost too much in various ways. Now, considering costs means going beyond what is supposedly self-evident.

Sixth comes Blanshard's doctrine of "rational will":

> that men have a common moral end which is the object of their rational will, that the state is a contrivance that they have worked out [that has evolved, Hayek would probably say] to help them realize that end, and that its authority over them rests on its being necessary for that end. If it is politically obligatory at times to obey a law that one regards as bad, that is because the state could not be run at all if the citizens could pick and choose which laws they would obey. Ultimately, therefore, political obligation, even that of obeying a morally bad law, *is* a moral obligation; and when, as occasionally happens, it become [sic] a duty to disobey, the ground is still the same. [1961, p. 395]

Four propositions elaborate this doctrine:

> First we can distinguish within our own minds between the end of our actual or immediate will, and the end of our rational will, which is what on reflection would commend itself as the greatest good. Secondly, this rational end is the same for all men. Thirdly, this end, because a common end, is the basis of our rights against each other. Fourthly, the justification of the state, and its true office, lie in furthering the realization of this end. . . . [T]he theory of a rational will provides a natural and intelligible ground both for obedience in normal cases and for disobedience in abnormal cases. [Pp. 395, 402]

Briefly interpreted, Blanshard's rational-will doctrine says that the obligation to support government is binding because—and to the extent that—it serves social cooperation. The obligation to support rather than subvert social cooperation rests, in turn, on ordinary ethical precepts. (I do not maintain, however, that Blanshard would himself accept the utilitarian label.)

We have no need for contractarian fictions. I might well obey the laws of an absolute monarchy, and even consider such obedience in the general interest and morally obligatory, while disapproving of that government's nondemocratic character and of some of its actions.

A Utilitarian Conception of Government

A restatement is worth attempting. We go along with the existing form of government and generally obey its laws because, first, we have no real alternative. For an individual, revolt would be fruitless and moving abroad too costly. (Besides, where would we go?) We individuals have not agreed even to having a government at all, much less to the particular constitution in force. We have not even been asked whether we agree (and asking us now would be a mockery). Rather, we of the current generation find ourselves living under a form of government and under laws that have evolved over time without our individually having had any effective say. Government and laws are *not* primarily results of an organized and deliberate process of collective decisionmaking, certainly not of one in which we the living have taken decisive part.

A second reason for acquiescence is that we find the existing system preferable to general lawlessness. Peace and security and a stable legal framework serve social cooperation and thus happiness. We individuals benefit from others' abiding by the law and feel that we in turn should do the same. We feel that it would be morally wrong to make exceptions in our own favor at others's expense. In self-defense we apply force against criminals who flout such of the moral code as has been reinforced by law.

The most—which is perhaps too much—that can be said for a social-contract theory is that most of us abide by the law and refrain from unconstitutional subversion of existing government in expectation or in consideration of others' doing the same. But this very nebulous contract, if it is a contract at all, is of the same sort as the one in accordance with which we generally observe ordinary ethical precepts. We ordinarily show some consideration for other people and their rights because we expect them to show similar consideration for us and because behaving with this consistency and decency serves our own self-esteem. Considerateness for each other yields gains from trade.

It would really be reaching, however, to interpret this sort of implicit trading as a social contract, and particularly as a contract whereby each of us has consented to the existing constitution and is thus deemed to have consented to government decisions made in accordance with the constitution. It is an exaggeration to call the government's laws and actions the result of collective decisionmaking in any literal sense of the term. Let's face it: government decisions are made by government officials (and the composite of those decisions undergoes some unintended drift over time); we ordinary citizens are not the government.

Under democracy, it is true, we have some influence on those decisions through voting, through helping shape public opinion, and thus through influencing what decisions public officials will consider in their own interest. But our control over government is less precise and effective than our control over economic activity through our "voting in the marketplace." Some analogy does hold between political and economic decisions, but we should not delude ourselves about how closely it holds.

Is the state a product of its citizens' voluntary consent, a mechanism voluntarily established to attend to their common concerns? Nonsense. I have no choice about being subjected to its laws. True enough, I am glad that the state exists; I prefer it to anarchy; but the state is there whether I want it or not. My welcoming certain arrangements does not mean that they are not compulsory. I am glad to have seat belts in my car and would probably have bought them willingly if I had had a free choice, but the fact remains that I did not have a free choice and that the belts were installed under compulsion of law.

Far from the state's being a voluntary arrangement, then, its essence is compulsion. It relies as a last resort on its power to seize goods and persons, to imprison, and to execute. If obedience to government is not compulsory, then what is? What does the word "compulsory" mean? What happens to the distinction between the voluntary and the compulsory?

To say this is not to glorify the compulsory aspects of government. I concede their necessity only with regret. I want to keep them tightly restrained, as the cause of human liberty requires. One serves that cause poorly if one deludes oneself into thinking that government embodies free exchange and that compliance with its orders is voluntary. Hard-headedness or tough-mindedness better serves one's values.

While libertarians want to extend the voluntary aspects of society and government, they should not delude themselves about reality and the human condition. Society and government are not and cannot be the results of a social contract. Their justification rests on other considerations.

The key element in the case for democracy, as I see it, is that democracy lessens the necessity or desirability of violent rebellion. It makes the alternative, discussion, relevant. If a policy or a law really is oppressively bad, citizens and their political representatives may come to understand why and may change it peacefully. This case for democracy is a far cry from asserting that all decisions made under democratic government are therefore made in accordance with each citizen's will, or his real will, or are to be "considered" as having been so made. We need not appeal to any fiction about unanimous constitution agreement to waive unanimous agreement on specific issues.

Tacit Utilitarianism Among Rights Theorists

Even several rights theorists who disavow utilitarianism do tacitly employ a version similar to the one recommended in this paper. I ask them to conduct a mental experiment. Suppose, just for the sake of argument, it could be demonstrated that insistence on the inviolability of human rights as they conceive of them would lead to general misery, whereas a pragmatic policy of respecting rights or not as conditions seemed to recommend would lead to general happiness. Would those theorists still insist on the inviolability of rights as the supreme goal to be upheld even at the cost of prevalent human misery?

Perhaps they would reply that this is a preposterous supposition and that respect for rights promotes human fulfillment and happiness, whereas a pragmatic attitude toward rights is an obstacle. Well, I think so too. But unless the rights advocates do answer "yes" to the question, insisting on rights even at the cost of general misery, they are taking a broadly utilitarian stance. If they answer as I think they must, they are insisting on rights because of the good consequences of upholding them and the bad consequences of disregarding or overriding them. At the back of their minds, at least, they must have some notion of a workable social order as an indispensable means to happiness.

Why do I care about the word "utilitarian"? Why am I anxious to pin that label onto everyone? Well, I do not care about the word as such. (And I do distinguish between versions of utilitarianism, although I do not find the supposed distinction between utilitarianism and consequentialism of much importance.) However, when a doctrine that plausibly and in accord with established usage bears the label "utilitarian" comes under attack, it serves clear thinking and communication to defend that doctrine under its own name rather than cast about for a new one. Playing the latter game is like trying to defend capitalism by inventing a new name for it. The game seems to admit that the doctrine or system defended really is so odious that it must be referred to only by euphemisms.

Robert Nozick, who avowedly just postulates rights without developing an argument for them, provides an example of tacit utilitarianism in the way he handles the question of blackmail. Murray Rothbard, another rights theorist, had put blackmail on a par with any other economic transaction; it would not be illegal in a free society (1962, vol. 1, p. 443, n.49). Nozick counters that blackmail is wrong, akin to the protection racket. He takes a step toward assessing its effect on the character of society by noting that it, like the racket, is an unproductive activity, whereas bona fide protective services are productive. (Nozick 1974, pp. 85–86. Rothbard 1977, pp. 53–55, rejects Nozick's distinction, but with arguments that strike me as feeble, even though they too are in part, unavowedly, utilitarian.)

Tacit utilitarianism creeps into the discussion of risky activities also. Suppose that your neighbor handles explosives recklessly or plays Russian roulette with a cannon mounted on a turntable. Even if he has liability insurance, he harms you, probably by raising your own insurance rates, lowering the value of your property, and striking fear and apprehension into you. You might plausibly argue that your neighbor is infringing on your rights, even though no explosion or cannon ball happens to damage your house. Rothbard appears to brush aside such problems by asking, in effect: If "fear" of others' "risky" activities is allowed to justify action against them, won't *any* tyranny become justified? What about the greater risk of having a state empowered to control activities it deems risky? (Rothbard 1977, especially pp. 48–50. The particular example used here is mine, not Rothbard's or Nozick's, but it suits the general tenor of their discussion.)

Nozick (1974, pp. 74–75, 78), on the other hand, candidly recognizes that "Actions that risk crossing another's boundary pose serious problems for a natural-rights position. . . . Imposing how slight a probability of a harm that violates someone's rights also violates his rights? . . . It is difficult to imagine a principled way in which the natural-rights tradition can draw the line to fix which probabilities impose unacceptably great risks upon others." Many kinds of actions do impose some degree of risk on others. A society that prohibited them all unless the actors had adequate means or adequate insurance to pay for possible harm would "ill fit a picture of a free society as one embodying a presumption in favor of liberty, under which people permissibly could perform actions so long as they didn't harm others in specified ways."—Again, the good-society approach, utilitarianism!

Other utilitarian strands are evident in Nozick's book. His flexibility about property rights is an example. He supposes that a natural disaster destroys the entire supply of water except one man's, which is sufficient for everyone. Under these circumstances, other persons may take the water or at least are not obliged to pay whatever exorbitant price its owner may demand. Nozick appeals to the Lockean proviso that one man's appropriation of a resource is justified only if it leaves enough and as good of that resource for others. He is not, in his own view, saying that recognized property rights may be overridden. Instead, "Considerations internal to the theory of property itself, to its theory of acquisition and appropriation, provide the means for handling such cases." (1974, pp. 180–81. Nozick adds another example: "Similarly, an owner's property right in the only island in an area does not allow him to order a castaway from a shipwreck off his island as a trespasser, for this would violate the Lockean proviso.")

It is mere word play, however, to wonder whether rights are overridden or are defined in the first place to be appropriate even for the catastrophe case. Nozick misapplies the Lockean proviso anyway, since it concerns someone's original acquisition of property, not its retention in the face of changed circumstances of other persons (Hodson 1977, especially pp. 221, 224–27).

Nozick tacitly appeals to utilitarian considerations in framing his conception of property rights in a sufficiently complicated and flexible way to allow the actions that intuition and utility would suggest in the catastrophe case. He would have us permit acts that threaten to cross our boundaries—loosely speaking, violate our property rights—when certain conditions are satisfied, including those of the case in which the benefits in harm prevented or good produced far outweigh the costs of fully compensating the person whose boundaries are crossed. (Rabinowitz 1977, p. 93. Lawrence A. Scaff 1977, p. 202, looks behind Nozick's assertion that moral theory has priority in political discussions and finds his language of moral theory consisting of "economic terms, calculations, categories, and assumptions. Moral discourse is suffused with cost-benefit analysis. Thus, even in the realm of morality, all values carry a price tag.")

Nozick (p. 79) adduces similar considerations in recommending cost-benefit analysis and the test of compensation (actual or merely potential?) in decisions on which polluting activities to forbid and which to permit. He recognizes (p. 182) that he cannot derive a definite position on patents from considerations of rights alone. Although a patent does not deprive others of what would not exist if not for the inventor's work, knowledge of the patented invention does tend to discourage independent efforts to reinvent it. "Yet . . . in the absence of the original invention, sometime later someone else would have come up with it. This suggests placing a time limit on patents, as a rough rule of thumb to approximate how long it would have taken, in the absence of knowledge of the invention, for independent discovery."

Tibor Machan, avowedly a rights theorist, is another tacit utilitarian. Instead of simply postulating or intuiting rights, he inquires into the political principles of a good society—good for man's pursuit of happiness or perhaps excellence, given his nature and his character as a moral agent. Machan seeks to demonstrate that since "each person is responsible to achieve his own happiness, that society that is suited for him is one in which his individual liberty is fully secured. . . . [I]t is only in such a free society that the moral agency, the freedom and the dignity, of each person can be respected. Thus only in that kind of community can the moral life flourish" (1975, p. 100).

Machan asks why someone who has the option of either taking from another person or producing and trading on his own should support the institution of private property. He suggests essentially utilitarian answers, reasons why respecting the institution of ownership is generally advantageous even for the person in question. He concludes "that ownership is a morally appropriate institution for human beings in general. . . . [T]o rely on *his* own work (and/or trade, creativity, ingenuity, etc.) is better for the person than to live off the work of others" by stealing or confiscating (pp. 133–34). Finally, Machan's approving characterization of Ayn Rand's doctrine is tacitly utilitarian. Rand, he says, defends "capitalism as morally right because human beings can work (trade, create, risk) for their own good only when and where it prevails." She advocates it as "a system that is good for human beings, morally good for them, to choose for themselves" (Machan 1975, p. 136).

Even Murray Rothbard has at least once (1973, pp. 23–25) lapsed into a tacit utilitarianism, seeing it as "vitally necessary for each man's survival and prosperity that he be free to learn, choose, develop his faculties, and act upon his knowledge and values. . . . [T]o interfere with and cripple this process by using violence goes profoundly against what is necessary by man's nature for his life and prosperity."

Tacit Utilitarianism Among Contractarians

In the contractarian camp, John Rawls in effect says we should ask: "If a group of ideally rational beings came together in order to pick rules to govern their mutual relations, which rules would they be compelled (by the power of their rationality) to pick?" (restatement by Murphy 1977, p. 233). Well, what do those beings rationally take into account? Facts of reality and applicable economic and other theories, presumably, together with a value judgment in favor of happiness, especially their own. Rawls assumes that the parties negotiating in the original position already accept a "thin theory of the good," according to which "liberty and opportunity, income and wealth, and above all self-respect are primary goods," goods conducive to happiness for persons almost regardless of their specific personality traits and plans of life. This list of primary goods "is one of the premises from which the choice of the principles of right is derived" (Rawls 1971, pp. 397, 433–34, and *passim*).

When Rawls considers what principles would help make primary goods available and so serve persons' pursuit of their life plans, he is appealing to broadly utilitarian considerations. The same is true when he rhapsodizes over the benefits of public commitment to avowed principles of justice: "... deliberate injustice invites submission or resistance. Submission arouses the contempt of those who perpetuate injustice and confirms their intention, whereas resistance cuts the ties of community" (p. 384). In discussing the problem of envy, Rawls notes the advantages of having a plurality of associations and many noncomparing groups. His "principles of justice are not likely to arouse ... envy ... to a troublesome extent.... What a social system must not do clearly is to encourage propensities and aspirations it is bound to repress and disappoint. So long as the pattern of special psychologies elicited by society either supports its arrangements or can be reasonably accommodated by them, there is no need to reconsider the choice of a conception of justice.... [T]he principles of justice as fairness pass this test" (1971, pp. 536–37, 541; Rawls's discussion of envy covers pages 530–41). Furthermore, Rawls's whole method of reflective equilibrium—testing tentative principles by how they are likely to work out in practice and adjusting both principles and judgments about particular cases to achieve consistency between them—is a kind of utilitarianism.

Rawls himself rejects this label. He recognizes that the parties in the original position might adopt some form of utility principle in defining the principles of social cooperation. Still, he says, it would be "a mistake to call these principles—and the theory in which they appear—utilitarian.... In fact, the case for the principles of justice is strengthened if they would be chosen under different motivation assumptions." Contract theory could eventually lead "to a deeper and more roundabout justification of utilitarianism" (1971, pp. 181–82). In saying so, Rawls is forgetting that his notion of a contract negotiated in an original position is utter fiction.

Buchanan's wing of contractarianism has already been described sufficiently to suggest how it is tacitly utilitarian. According to its tenets, an economist is entitled to recommend a policy only tentatively, only as a hypothesis that it is in accord with a unanimously made contract, or that it conceptually commands agreement, or that it could command agreement, presumably after a sufficient amount of sufficiently enlightening public discussion. Well, on what basis could the economist expect or hope for the necessary degree of agreement? The probable effects of the contemplated policy must surely figure prominently in the answer.

Contractarians like Buchanan distinguish between propounding hypotheses about what policies could ideally command agreement and recommending policies because they are expected to enhance social cooperation and so serve human happiness. The distinction seems operationally empty to me.

Conclusion

The pure doctrine of natural or human rights cuts analysis short either by merely postulating rights as axioms or by questionably deriving rights from supposedly axiomatic propositions that in fact require further examination themselves. In truth we cannot infer one infallibly best set of institutions and policies from one or more first principles whose implications are guaranteed never to clash. It is a "great illusion" in political philosophy to seek "solutions to insoluble problems. . . . [T]here is more than one basic principle that appeals to moral sense and for which good argument can be made. . . . We live in a morally messy world. But it is the one we are stuck with" (Gordon 1976, p. 589).

Contractarianism rests on farfetched fictions. Or if it does not exactly rest on them, its rhetoric does abound in them; and if it is stripped of its fictions and translated into straightforward language, contractarianism turns out to be not much different from a form of utilitarianism.

We can hardly make progress in social philosophy or policy analysis by adopting *fictions* as our first principles. While wishing to enhance the voluntary and market-like aspects of government, for example, we must not blind ourselves to its essentially coercive character. Instead of beguiling ourselves with attractive myths, we can better serve our fundamental values by trying to compare alternative sets of institutions, alternative big pictures, avoiding excessively narrow and short-run focus. Investigation, analysis, and discussion of the features and probable consequences of contemplated institutions and policies all are indispensable aspects of the search for agreement—assuming, for the sake of argument, that agreement were the touchstone of policy. Actually, agreement itself cannot form the decisively appealing substance of a state of affairs capable of commanding it.

Discussion in search of agreement relies ideally on investigation, reasoning, and checking and comparison—the ordinary scientific process. This process is poles apart from dictatorship and from appeals to infallible insight. In this process, we can communicate better and guard better against misunderstandings if we employ straightforward language.

Conceivably the rights doctrine and contractarianism are not the only alternatives to the approach to policy espousal that I recommend.

If so, those other alternatives deserve further investigation and discussion. Meanwhile, I submit, a rules-utilitarianism or indirect utilitarianism—in other words, a good-society/comparative-institutions approach—turns out to be the only one that stands up under critical inspection.

References

Adler, Mortimer J. *The Time of Our Lives*. New York: Holt, Rinehart, and Winston, 1970.

Alexander, Sidney S. "Human Values and Economists' Values." In *Human Values and Economic Policy*. Edited by Sidney Hook. New York: New York University Press, 1967.

Bentham, Jeremy. "A Critical Examination of the Declaration of Rights." Reprinted in *Bentham's Political Thought*, pp. 257–90. Edited by Bhikhu Parekh. New York: Barnes & Noble, 1973.

Blanshard, Brand. *Reason and Goodness*. New York: Macmillan, 1961.

Broome, John. "Choice and Value in Economics." *Oxford Economic Papers*, n.s. 30 (November 1978): 313–33.

Buchanan, James M. *Freedom in Constitutional Contract*. College Station: Texas A & M University Press, 1977.

Buchanan, James M. *The Limits of Liberty*. Chicago: University of Chicago Press, 1975a.

Buchanan, James M. "Natural and Artifactual Man." Manuscript.

Buchanan, James M. "Sources of Opposition to Constitutional Reform." In *Constitutional Economics*. Edited by Richard B. McKenzie. Lexington, Mass.: Lexington Books, 1984.

Buchanan, James M. "Utopia, the Minimal State, and Entitlement" (a review article on Robert Nozick, *Anarchy, State, and Utopia*). *Public Choice* 23 (Fall 1975b): 121–26.

Buchanan, James M. "Who Should Distribute What in a Federal System." In *Redistribution Through Public Choice*. Edited by H. Hochman and G. Peterson. New York: Columbia University Press, 1974.

Buchanan, James M. "Why Does Justice Have Value?" Manuscript.

Buchanan, James M., and Tullock, Gordon. *The Calculus of Consent*. Ann Arbor: University of Michigan Press, 1962.

Davis, Harry R. "Toward Justifying Democracy." *The Key Reporter* 33 (Winter 1967/68): 2–4, 8.

Edgeworth, Francis Y. *Mathematical Psychics*. London: Kegan Paul, 1881; photoreprint, New York: Kelley, 1961.

Edwards, Paul. *The Logic of Moral Discourse*. New York: Free Press, 1965.

Fletcher, Joseph. *Situation Ethics*. 1st British ed. London: SCM Press, 1966 (fourth impression, 1974).

Gewirth, Alan. *Reason and Morality*. Chicago: University of Chicago Press, 1978.

Gordon, Scott. "The New Contractarians." *Journal of Political Economy* 84 (June 1976): 573–90.

Gray, John N. "Social Contract, Community and Ideology." In *Democracy, Consensus & Social Contract*. Edited by Pierre Birnbaum, Jack Lively, and Geraint Parry. London and Beverly Hills: Sage Publications, 1978.

Gray, John N. *Mill on Liberty: A Defence*. London and Boston: Routledge & Kegan Paul, 1983.

Grinder, Walter. "Cross Currents." *Libertarian Review*, May 1978, pp. 9–10.

Halévy, Élie. *The Growth of Philosophic Radicalism*. Trans. by Mary Morris. Boston: Beacon Press, 1955.

Hare, R. M. "Rawls' Theory of Justice." In *Reading Rawls*. Edited by Norman Daniels. New York: Basic Books, n.d. (but apparently 1974).

Harsanyi, John C. *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht and Boston: D. Reidel Publishing Company, 1976.

Harsanyi, John C. "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (August 1955). Reprinted as selection #10 in *Economic Justice*, pp. 266–85. Edited by E. S. Phelps. Baltimore: Penguin Books, 1973.

Hayek, Friedrich A. *The Constitution of Liberty*. Chicago: University of Chicago Press, 1960.

Hayek, Friedrich A. *Freiburger Studien*. Tübingen: Mohr (Siebeck), 1969.

Hayek, Friedrich A. *Law, Legislation and Liberty*. Vols. 1–3. Chicago: University of Chicago Press, 1973, 1976, 1979.

Hayek, Friedrich A. *New Studies in Philosophy, Politics, Economics, and the History of Ideas*. Chicago: University of Chicago Press, 1978.

Hayek, Friedrich A. *Studies in Philosophy, Politics and Economics*. Chicago: University of Chicago Press, 1967.

Hazlitt, Henry. *The Foundations of Morality*. Princeton: Van Nostrand, 1964.

Hazlitt, Henry. "An LR Interview." *Libertarian Review*, November 1978, pp. 22–23.

Hobbes, Thomas. *Leviathan*. In *Great Books of the Western World*, vol. 23. Chicago: Encyclopaedia Britannica, 1952.

Hodson, John D. "Nozick, Libertarianism, and Rights." *Arizona Law Review* 19, no. 1, 1977, pp. 212–27.

Holcombe, Randall B. *Public Finance and the Political Process*. Carbondale and Edwardsville: Southern Illinois University Press, 1983.

Hume, David. "Of the Original Contract." Reprinted in *Hume's Ethical Writings*. Edited by A. MacIntyre. New York: Collier Books, 1965.

Kemp, J. *Ethical Naturalism: Hobbes and Hume*. London: Macmillan, 1970.

Machan, Tibor. *Human Rights and Human Liberties*. Chicago: Nelson Hall, 1975.

Mack, Eric. "An Outline of a Theory of Rights." Manuscript, 1978.

McCloskey, H. J. *Meta-Ethics and Normative Ethics*. The Hague: Martinus Nijhoff, 1969.

Meyer, Frank S. *In Defense of Freedom*. Chicago: Regnery, 1962.

Mill, John Stuart. *Utilitarianism*. In *Selected Writings*. Edited by M. Cowling. New York: New American Library, 1968.

Von Mises, Ludwig. *Human Action*. New Haven: Yale University Press, 1949.

Von Mises, Ludwig. *Theory and History*. New Haven: Yale University Press, 1957.

Murphy, Jeffrie G. "Rights and Borderline Cases." *Arizona Law Review* 19, no. 1, 1977, pp. 228–41.

Nozick, Robert. *Anarchy, State, and Utopia.* New York: Basic Books, 1974.

Paul, Jeffrey. "On the Foundations of Natural Rights." Manuscript, 1978.

Peirce, Charles S. "The Scientific Attitude and Fallibilism." In *Philosophical Writings of Peirce,* pp. 42–59. Edited by Justus Buchler. New York: Dover, 1955.

Pettit, Philip. *Judging Justice.* London and Boston: Routledge & Kegan Paul, 1980.

Philbrook, Clarence. " 'Realism' in Policy Espousal." *American Economic Review* 43 (December 1953): 846–59.

Prescott, Peter S. Review of Thornton Wilder, *Theophilus North. Newsweek,* 22 October 1973, p. 128b.

Rabinowitz, Joshua T. "Emergent Problems and Optimal Solutions." *Arizona Law Review* 19, no. 1, 1977, pp. 61–157.

Rawls, John. *A Theory of Justice.* Cambridge: The Belknap Press of Harvard University Press, 1971.

Rawls, John. "Two Concepts of Rules." *Philosophical Review,* 1955. Reprinted in *Contemporary Utilitarianism,* pp. 59–98. Edited by M. D. Bayles. Garden City: Anchor Books, 1968.

Rothbard, Murray. *The Ethics of Liberty.* Atlantic Highlands, N.J.: Humanities Press, 1982.

Rothbard, Murray. *For a New Liberty.* New York: Macmillan, 1973.

Rothbard, Murray. *Man, Economy, and State.* 2 vols. Princeton: Van Nostrand, 1962.

Rothbard, Murray. "Robert Nozick and the Immaculate Conception of the State." *Journal of Libertarian Studies* 1 (Winter 1977).

Samuels, Warren J. "The Myths of Liberty and the Realities of the Corporate State: A Review Article." *Journal of Economic Issues* 10 (December 1976): 923–42.

Scaff, Lawrence A. "How Not to Do Political Theory: Nozick's Apology for the Minimal State." *Arizona Law Review* 19, no. 1, 1977, pp. 193–211.

Schumpeter, Joseph A. *History of Economic Analysis.* New York: Oxford University Press, 1954.

Sen, Amartya K. *Collective Choice and Social Welfare.* San Francisco: Holden-Day, 1970.

Spencer, Herbert. *The Principles of Ethics.* 2 vols. Indianapolis: Liberty Classics, 1978 (originally published in 1897, and parts earlier).

Vanberg, Viktor. "Liberty, Efficiency and Agreement: The Normative Element in Libertarian and Contractarian Social Philosophy." Manuscript, George Mason University, n.d. (but 1984 or 1985).

Wiener, Philip P. "Charles Sanders Peirce." *International Encyclopedia of the Social Sciences*, vol. 11, pp. 511–13. 1968.

Yeager, Leland B. "Pareto Optimality in Policy Espousal." *Journal of Libertarian Studies* 2 (Fall 1978): 199–216.

Yeager, Leland B. "Utility, Rights, and Contract: Some Reflections on Hayek's Work." In *The Political Economy of Freedom: Essays in Honor of F. A. Hayek.* Edited by Kurt R. Leube and Albert H. Zlabinger. Munich and Vienna: Philosophia Verlag, 1984/85.

LIFE IN THE GULAG: A PROPERTY RIGHTS PERSPECTIVE

Gary M. Anderson and Robert D. Tollison

> No, it isn't done especially to torture people. A sentenced prisoner is a laboring soldier of socialism, so why should he be tortured? They need him for construction work.
>
> —Aleksandr I. Solzhenitsyn

Introduction

One of the most interesting applications of modern economic theory has been to the analysis of slavery. This work began in the pathbreaking article by Conrad and Meyer (1958), and was extended and revised by Fogel and Engerman (1974). This analysis, including a growing body of further extensions and critiques, focuses on the case of slavery in the antebellum South.

In this paper we show that the analysis of slavery based on the economic theory of property rights can be extended beyond the antebellum South to a more recent institutional phenomena—the large-scale use of political prisoners as forced labor in modern socialist states. These cases are amenable to a straightforward application of modern property rights theory, which offers a rational explanation for key differences between these examples of forced labor and slavery in the antebellum South.

The Plantation

One of the key insights of the modern economic literature on slavery is the importance of a highly developed, efficient capital

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

Gary Anderson is a graduate student in economics at the Center for Study of Public Choice, George Mason University, and Robert Tollison is Professor of Economics and Director of the Center for Study of Public Choice. The authors are grateful to Lee Alston and James Dorn for comments on an earlier version of this paper. The usual caveat applies.

market for slaves in the slave-holding region of the United States prior to the Civil War. This market determined prices for slaves based on the discounted present value of the future income streams expected to result from their labor services, as in the case of any other long-lived capital asset. The slave capital market existed in two major forms, a rental and a purchase market. Rentals were usually for periods ranging from a day to a year, and served as a transactions cost-minimizing device allowing the use of slave labor for temporary jobs with a flexibility similar to that of nonslave wage labor. The sale of slaves on organized slave markets is estimated to have averaged 50,000 per year, implying that on average 1 slaveholder out of every 22 sold a slave in a given year (Fogel and Engerman 1974, p. 53).

The point we wish to emphasize here is that the expected future value of a slave's labor services, which were a function of the slave's health, fitness, education, training, and other characteristics (and to a degree represented levels of discretionary investment by the slave owner), was reflected in the price of the slave determined on a capital market. Slave owners were not restrained by law from neglecting, injuring, or otherwise abusing their slaves. However, they were restrained by economic incentives from reducing the capital value of their slaves as assets.

In fact, there is evidence that rational maximizing behavior on the part of slave owners led to effective maintenance of the health and physical welfare of slaves. Slaves were not worked to death at early ages, but treated well enough so that their average life expectancy at birth was evidently similar to that of whites (Fogel and Engerman 1974, p. 126). Childbearing by slaves was encouraged, and the intentional breakup of slave families was usually avoided (Fogel and Engerman 1974, p. 142). Available evidence suggests that slave diets were nutritionally balanced (in fact, probably better balanced on average than the diets of freed slaves in the South in the later 19th century; Fogel and Engerman 1974, p. 113). In short, while antebellum slavery was an institution based on coercion and exploitation, slave owners had economic incentives to maintain the physical requirements for the basic health and welfare of their slaves.[1]

---

[1] It might be argued that legal protection of slaves in the South contributed an additional constraint on the mistreatment of slaves by their owners. In fact, 10 Southern states made it a crime to mistreat a slave, and the law of 6 states required the master to provide suitable food and clothing for his slaves. However, these laws were rarely enforced. No slave could legally testify against his master, and in some states a black could not testify against a white man at all. Hence, these laws were irrelevant to the master's decision concerning the optimal level of investment in slave maintenance. Interestingly, states commonly compensated masters if slaves were sentenced to capital punishment for crimes. For example, an 1824 law in Alabama entitled slaveowners to claim one-half of the assessed value of slaves put to death; in 1839, Virginia appropriated $12,000 to pay owners of executed slaves compensation. For a discussion of these matters, see Friedman (1974, pp. 198–200).

There was another feature of antebellum slavery based on private property rights with important consequences for the well-being of slaves. Owners, as full residual claimants in both the capital value and output derived from slave labor, were motivated by economic incentives to maximize slave productivity. Critics of the economic efficiency of slave labor have tended to argue that slaves, unlike free laborers, were not residual claimants in the value of their human capital stock, and hence did not bear the cost of shirking except in the form of punishment. Given that punishment is the only form of control that masters have over slave labor, slaves must be externally monitored during labor. Free labor tends to require less costly monitoring and to be partially self-monitoring for reasons involving residual claimancy.

However, Fogel and Engerman (1974, pp. 147–49) demonstrate that Southern slave owners did not rely exclusively on the lash for supplying slave worker motivation, but supplemented physical coercion with economic incentives in the form of output-based rewards. The use of rewards for motivating slaves was widespread. Even profit-sharing arrangements between slaves and their masters were occasionally employed. Further, rewards in the form of occupational advance were employed. Although the occupational opportunities of slaves were restricted by comparison with those of white nonslaves, advance within the occupational hierarchy of the plantation based on individual effort was possible. Taken together, these forms of reward constituted an incentive structure which partially replaced monitoring and control by the slave owner with self-monitoring by slaves who were made residual claimants in the production process.[2]

The Gulag

Contrary to popular belief, slavery in the antebellum South is not the last significant example of a system of slave labor in recent history. The forced labor camps in various socialist states in this century constitute a well-documented but poorly understood modern system of slave labor to which the same basic analysis as that outlined above applies. Obviously, there are important differences between the system of forced labor in the Soviet Union and that of slavery in the antebellum South. Our purpose is to emphasize the economic importance of some of these differences.

---

[2] It should be noted that Fogel and Engerman's work on antebellum slavery has generated considerable controversy. Several critics of *Time on the Cross* published a book-length detailed critique (David, et al. 1976), based on primary source material. Of particular relevance here are Sutch's article, which argues that Fogel and Engerman exaggerated the nutritional content of the slave diet (David, et al., pp. 231–301), and Gutman and Sutch's article, which maintains that whipping and other forms of corporal punishment of slaves were significantly more important than Fogel and Engerman claimed (David, et al., pp. 53–93). However, despite the controversy, both sides agree that slaveowners tended to preserve the value of slaves as capital assets and that this behavior was economically motivated.

Unlike slavery in the antebellum South, individuals committed to forced labor in the USSR are convicted criminals and political prisoners. Therefore, forced labor camps, unlike slave plantations, are designed in part to punish their inmates. However, the forced labor camps are also productive enterprises generating a nontrivial proportion of the output of the Soviet economy.

According to one recent estimate, the total number of *zeks* (inmates in forced labor camps) in the USSR at any given time is about five million adults (Orlov 1979, p. 68). Dallin and Nicolaevsky (1947, p. 15) claim that the camp population was between 7 to 12 million prisoners in any given year throughout the 1930s and 1940s. Following Stalin's death in 1953, the camp population fell by about two-thirds. Since 1957, however, it has trended upward (Conquest 1971, p. 21).

This forced labor pool is allocated to activities such as mining in remote and inhospitable regions (such as eastern Siberia where the bulk of the large labor camps are located). Forced labor has played a leading role in the economic development of resource-rich areas of the wilderness in Siberia and elsewhere in the USSR. This is not to imply that the use of forced labor in the Soviet Union is efficient relative to voluntary labor, but merely that forced labor camps have been responsible for producing significant economic output in absolute terms. Forced labor is a productive input and not only a form of punishment.

But the most relevant difference between slavery in the antebellum South and in the Soviet prison camps (as well as its counterparts in other communist countries) has been neglected in the large literature which has emerged about the Gulag. *This is the fact that the forced labor system of the Gulag is an example of slavery in the absence of well-defined and enforced property rights in slaves.*

Inmates in Soviet forced labor camps are not slaves in the strict sense because they do not represent private property. However, *zeks* are not protected under the law, but are under the control of the respective camp administration. These inmates are therefore not self-owners. But neither are they owned by camp administrators or any other particular individuals. They are not bought and sold and hence do not have capital values.

The standard textbook critique of economic efficiency under socialism emphasizes in part the incentive problems that result from the absence of well-defined and enforced private property rights to plant, equipment, and resources. Decision makers in the producer sector are not residual claimants, and thus lack economic motivation to make efficient decisions. Socialism tends to be characterized by the absence of efficient property rights arrangements that would internalize economic externalities.

Even slave labor under socialism is subject to these problems. Inmates in the Gulag may not be bought and sold by camp administrators; there is no capital market for slaves in the Soviet Union (just as there is no capital market for other productive assets). *Zeks* therefore have no market-determined capital value.

In the Gulag, as in other parts of the planned sector of the Soviet economy, decision makers are not residual claimants in the direct sense that entrepreneurs are in a market economy. Camp administrators (like factory managers) are typically assigned production quotas in terms of the relevant output (for example, X tons of gold processed per annum, or Y miles of highway constructed). They are evaluated and rewarded on the basis of the degree to which they fulfill planned output. Also, the bonuses the manager/administrator can earn as a reward for good performance are usually relatively small. For these reasons camp administrators can be considered to be residual claimants only in a loose, indirect sense. The incentive of camp administrators to maintain the productivity of slave labor tends to be quite limited relative to a system of slavery in which the slave owners have full claim to the value of slave productivity.

The abominable treatment of prisoners in the Gulag Archipelago has been discussed extensively, most especially by Solzhenitsyn (1974) in his trilogy of that title documenting conditions in the camps during the Stalin era and shortly thereafter. Solomon (1971, p. 103) estimates that 40 million prisoners died in forced labor camps prior to 1950; most sources agree that death rates in the forced labor camps have been consistently high. Solzhenitsyn documents the deplorable sanitary conditions, inattention to the health of inmates, and virtual starvation diets which appear to have reflected standard camp operating procedures.

The relative harshness of the Soviet system of forced labor by comparison with most historical instances of slavery is typically explained in the literature as reflecting the intentions of the regime. The Gulag is asserted to have been designed to produce a joint product, in the form of economic output and punishment. Conquest (1978, p. 124) explains the brutal conditions at the huge Kolyma camp complex as reflecting "one main truth," namely, that "in the minds of its creators and organizers the conscious purpose of Kolyma . . .

had become the production, with at least equal priority, of gold and death." In fact, the Soviet system is sometimes portrayed as linked to the Gulag to provide stability for the regime. Again quoting Conquest (1978, p. 231): "Kolyma—the threat and actuality of Kolyma—was the way the Soviet government imposed itself on its subjects."

While this line of reasoning seems plausible, it fails to explain the way that the Gulag Archipelago actually operates. Given serious labor shortages in Siberia, where the bulk of the camps are located, the inmate population represents valuable labor inputs. A rational (albeit despotic) regime would seek to conserve such resources rather than casually waste them. In the event that the cost of sending an additional prisoner to the camps was perceived to exceed that individual's expected marginal value product, it would be more efficient for the regime to arrest and murder the individual near his home. Shipping individuals thousands of miles to camps where their labor is expected to be submarginal in order to work them to death would be an extraordinarily inefficient method of execution.

The puzzling feature of the forced labor system is why the prisoners have in general been treated as if they represented a costless labor resource. Slaves in the antebellum South were, in effect, prisoners for life, and were often treated harshly, but they were in general maintained as valuable, productive assets. They were not worked to death.[3] Slaves were valuable capital assets, and their value increased steadily following the legal prohibition of the importation of slaves into the United States in 1807.[4]

But in the USSR during the Stalin era, slaves in the Gulag could be replaced at essentially zero cost by those managing forced labor camps. While it was true that the risk of incarceration in a forced labor camp tended to vary among different groups in the society, the potential slave population basically included the entire population.[5]

---

[3] While this was true of slavery in the United States, there is evidence that at times in other parts of the Western Hemisphere (particularly in various parts of the Caribbean) some slaves apparently *were* worked to death. We would argue that in most of these cases transactions costs prevented capital markets from functioning efficiently, and that our property rights analysis of the Gulag system would apply to these cases as well. For a detailed economic model designed to explain why material conditions may differ under slavery, see Fenoaltea (1984).

[4] However, increasing prices for slaves during the antebellum era were not solely the result of the 1807 import prohibition. As Fraginals, Klein, and Engerman (1983) have recently shown, prices for slaves in Brazil and Cuba rose markedly through the 1850s, even though slave importation was legally permitted in both countries.

[5] Conquest (1978, p. 228) argues that most prisoners in the Gulag were not common criminals or dissidents, but had been arrested and convicted on trumped-up charges. It is clear that at least during the Stalin era the supply of hard labor to the camps was not fixed but highly elastic. The camp population included large numbers of people

A temporary labor shortage in the gold and coal mines near Omsukchan in Siberia, say, could be remedied by arresting and transporting a sufficiently large group of able-bodied dissidents, or if actual dissidents were not available, whatever group was convenient for the NKVD to deport. In other words, the supply of labor for the forced labor camps of the Gulag appears to have been nearly perfectly elastic from the perspective of the managers of the camps.

Under the incentive structure within which forced labor camp administrators operate, coupled with a high elasticity of supply for replacement slaves, the extremely high death rates within the camps, as well as conditions which tended to hamper the health and productivity of the inmates, can be explained without resort to any presumption of irrational malice on the part of the administrators. Gulag administrators did not own the inmates, and could not claim their capital value. Hence, they had a very substantially reduced incentive relative to slave owners to invest scarce resources in the care and maintenance of these laborers.

We are not arguing that the wasteful use of slave labor in the Gulag was a *necessary* consequence of the absence of explicit legal private property rights in slaves. Different institutional arrangements short of full private property rights and organized slave markets might have reduced waste. Effective property rights do not necessarily require explicit legal definition and enforcement.[6] If camp administrators were rewarded by higher authority for efficiently managing the slaves under their control, this would be tantamount to a limited form of property right in slaves, and would have the tendency to mitigate the exceptional brutality of the Gulag. Even in the absence of an explicit capital market for slaves, slaves would have a shadow price (and consequently an implicit capital value), reflecting their value as inputs in production to fulfill planned output, for which camp administrators are directly rewarded. This shadow price would

---

who were apparently not guilty of crimes or even suspected of crimes against the regime. During World War II (and later), large numbers of minority nationalities were deported to the camps (Conquest 1971, pp. 20–21). The NKVD followed the Red Army into occupied countries "to dispatch back millions of slaves" (Tolstoy 1982, p. 11). One authority (Tolstoy 1982, p. 18) refers to the universal vulnerability of ordinary Soviet citizens (regardless of their behavior toward the authorities) by describing the USSR in the Stalin era as populated by "prisoners, former prisoners, and future prisoners." The same author also explains that although the casualty rate in the forced labor camps was very high, "the remaining 'free' population provided a near inexhaustible source of fresh human material" (Tolstoy 1982, pp. 10–11).

[6] For example, see Umbeck (1981) on mining camps, and Libecap (1978) on range rights, both interesting examples of the emergence of property rights systems in the absence of explicit legal enactment.

perform the same functions as an explicit market price, albeit less efficiently.

In fact, camp administrators are responsible for maintaining specified output quotas in whatever their camps produced. Unfortunately, the low replacement cost of slaves facing camp decision makers prevents effective property rights in slaves from emerging. In the short run, targeted output would be maximized by maximizing the intensity of labor inputs supplied by slaves. Such a strategy would be efficient from the perspective of a camp manager if the marginal cost of replacement slaves is zero (or very low). It is also likely that other resources which are substitutable for slave labor as inputs in the production process—equipment and machinery, better management, and more efficient camp administration—are significantly more costly to camp administrators than are new *zeks*. Under these circumstances, resources provided to the camps for operating purposes can be consumed directly by the administrators, while still achieving planned output targets, by working the "free" slave labor to death. In other words, this reallocation of resources represents a way by which administrators could spend the lives of slaves in exchange for enhanced welfare for themselves.[7]

Of course, even in the short run, achieving an optimal rate of output per slave may require some minimal level of investment in maintenance (for example, providing some limited shelter, nourishment, and rest), such that some especially strong and healthy individuals may actually survive for years. Also, assuming that the transactions costs associated with procuring slave replacements are positive (that is, to the camp management)—for example, if replacement *zeks* require some limited degree of training for their assigned tasks—the camp management will have an additional incentive to engage in minimal investment in maintenance. But unless these transactions costs are very high, more costly maintenance such as medical care would not be efficient. Available evidence suggests that these conditions characterized the forced labor camp system during the Stalin era.

Conclusion

It has been common for critics of the Gulag Archipelago, and similar systems of forced labor in other socialist states, to describe it as slavery. This description is inaccurate because inmates of the Gulag, unlike slaves, are not bought and sold. As a consequence, they have no market-determined capital value. But more to the point, *zeks* are not *owned,* but represent free goods from the perspective of their masters. In this context the high wastage rate of inmates of the Gulag by comparison with antebellum slavery is economically understandable.

---

[7] Although camp administrators have opportunities to employ Gulag slaves to provide themselves with personal consumption goods (servants, sexual services, home construction, and so on), such personal use of slaves is difficult to conceal from central authorities, who monitor the camps closely. It is therefore quite limited. On the personal services provided by *zeks* to camp administrators, see Solzhenitsyn (1974).

The only thing worse than an economic system based on slave labor is one that treats slave labor as a free good. The consumers outside the Gulag have also suffered as a result of this wasteful system. By the early 1940s, as much as 25 percent of total direct investment in the Soviet Union was in the forced labor sector (Fainsod 1953, pp. 385–87). The net costs to the economy as a whole of transferring millions of workers from the ordinary sector to the Gulag have been large.

However, the Soviet economy is obviously not directed by the demands of consumers but by the demands of the central authorities. The possibility should be admitted that the Gulag represents an efficient means of producing the output mix demanded by the Soviet authorities. The forced labor system has proved a reliable labor supply at basically zero wage rates under conditions where a volunteer labor force would be available only at high real wage rates. At the same time, the stability of the regime may have been significantly enhanced by the terror the Gulag provided. Whatever the objective function of the rulers in Moscow, it is clear that the property rights arrangements operative within the system encourage massive waste in terms of slave labor resources.

References

Conquest, Robert. *The Human Cost of Soviet Communism.* Washington, D.C.: U.S. Government Printing Office, 1971.

Conquest, Robert. *Kolyma: The Arctic Death Camps.* New York: Viking Press, 1978.

Conrad, Alfred H., and Meyer, John F. "The Economics of Slavery in the Antebellum South." *Journal of Political Economy* 66 (April 1958): 95–130.

Dallin, David J., and Nicolaevsky, Boris I. *Forced Labor in Soviet Russia.* Hartford, Conn.: Yale University Press, 1947.

Fainsod, Merle. *How Russia is Ruled.* Cambridge, Mass.: Harvard University Press, 1953.

Fenoaltea, Stefano. "Slavery and Supervision in Comparative Perspective: A Model." *Journal of Economic History* 44 (September 1984): 1–21.

Fogel, Robert William, and Engerman, Stanley L. *Time on the Cross: The Economics of Negro Slavery.* Boston: Little, Brown and Company, 1974.

Fraginals, Manual Moreno; Klein, Herbert S.; and Engerman, Stanley L. "The Level of Structure of Slave Prices on Cuban Plantations in the Mid-Nineteenth Century: Some Comparative Perspectives." *American Historical Review* 88 (December 1983): 1201–18.

Friedman, Lawrence M. *A History of American Law.* New York: Simon and Schuster, 1974.

Gutman, Herbert, and Sutch, Richard. "Sambo Makes Good, or Were Slaves Imbued with the Protestant Work Ethic?" In *Reckoning with Slavery: A Critical Study in the Quantitative History of Negro Slavery,* pp. 55–93. Edited by Paul A. David, et al. New York: Oxford University Press, 1976.

Libecap, Gary D. "Economic Variables and the Development of the Law: The Case of Western Mineral Rights." *Journal of Economic History* 38 (June 1978): 338–62.

Orlov, Yuri. "On Prisoners in Soviet Camps." *Survey* 24 (Spring 1979): 67–85.

Solomon, Michael. *Magadan.* New York: Auerbach, 1971.

Solzhenitsyn, Aleksandr I. *The Gulag Archipelago 1918–1956.* Vol. 1. New York: Harper and Row, 1974.

Sutch, Richard. "The Care and Feeding of Slaves." In *Reckoning with Slavery: A Critical Study in the Quantitative History of Negro Slavery,* pp. 55–93. Edited by Paul A. David, et al. New York: Oxford University Press, 1976.

Tolstoy, Nikolai. *Stalin's Secret War.* New York: Holt, Rhinehart and Winston, 1981.

Umbeck, John. "Might Makes Rights: A Theory of the Formation and Initial Distribution of Property Rights." *Economic Inquiry* 19 (January 1981): 38–59.

STATE EDUCATION: HAVE ECONOMISTS MADE A CASE?

Jack High

Introduction

In 1662 economist William Petty argued that education should be a "Publick Charge." Looking at 17th-century Britain, Petty (1899, p. 27) saw entirely too many preachers, lawyers, and doctors to suit him. Were government to control education, he said, it could reduce the number of parishes by half, it could cut the number of lawyers by an astounding 99 percent ("I cannot conceive how there should remain above one hundredth part of what now are . . ."), and it could extinguish that "infinite swarm of vain pretenders unto, and abusers of [medicine]." The state, by deciding how many should be educated for each profession, could eliminate waste and further economic progress.

Petty's general attitude (if not his disdain for particular professions) was typical of early economists, who felt that society would improve in various ways if the state would take an active hand in education (see Johnson 1964). It was not until Adam Smith's *Wealth of Nations*, which appeared in 1776, that economists changed their uncritical approval of state education and began to look on it as an activity with drawbacks as well as benefits.

Once Smith had convincingly demonstrated that, as a general rule, natural liberty and decentralized decision making would promote society's interests better than government direction, education had to be justified as an exception to this rule. Why would governmental

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Assistant Professor of Economics and Executive Director of the Center for the Study of Market Processes at George Mason University. He is indebted to Tyler Cowen, Dan Klein, George Smith, E. G. West, and an anonymous referee for helpful comments, and to the Fred C. Koch Foundation for financial support.

bureaucracy produce better education than freedom of educational choice and the discipline of market competition?

Economists have provided various answers to this question, but the question itself has profoundly influenced the way economists have analyzed state education. Since Smith the presumption has been that free choice and market competition are preferable to coercion and government decree. Exceptions to this presumption must be explicitly stated and argued.

The thesis of this paper is that in the 200 years since Smith, economists have not made a compelling case for excepting education from free choice and market competition. Indeed, there are cogent reasons, many of which have been advanced by the same economists who favor some state involvement in education, for leaving education to the voluntary choices of the marketplace.

Although the arguments offered here are placed in the context of an intellectual exchange that extends back 200 years, the conclusions of this paper have a relevance for modern debate and policy, in that the economic dispute over proper government policy toward education continues to the present (see Friedman 1955; Vaizey 1962; West 1965, 1976; Friedman and Friedman 1980). These conclusions also bear on the historical debate sparked by Bailyn (1960), Katz (1968, 1971), and other revisionist historians.[1]

Also, during the last two decades, criticism of public schools has been harsh, and it has come from many points along the political spectrum. Most of the reform measures, however, have merely advocated changes in the operation or finance of public schools, often with the plea that more tax money be pumped into the system. (President Reagan's call for merit pay for teachers is an example.) The analysis in this paper implies that reform should aim at eliminating tax funds for, and removing state control over, education. Despite the relevance of the paper's thesis for current problems in education, the focus here is directed on the general arguments that economists have advanced in favor of government involvement in the education market.

Classifying the Arguments

Economists' arguments favoring state education fall into two broad categories—ethical and economic.[2] Ethical arguments maintain that state education can foster desirable moral traits among the populace; economic arguments maintain that state education increases efficiency or output. Within the ethical category, economists have argued that public education will improve the moral standards of the poor, foster values essential to a good society, reduce crime, and bring about social equality. Within the economic category, economists have claimed that public education will increase productivity and internalize external benefits. This last argument is especially prominent in modern economics.

---

[1] George Smith (1981) documents a number of similarities between advocates of voluntary education and the revisionist historians.

[2] Modern economics tries to expunge any ethical content from its theories, to deal with "positive" rather than "normative" propositions. By and large this was not true of the classical economists. Ethical considerations therefore constitute a large part of economists' arguments on government education.

These categories are not mutually exclusive (promoting proper values, for example, can increase economic production), and some of the arguments do not fit neatly into one category or the other. Nevertheless, this two-category classification does systematize the long list of arguments that economists have made over the course of two centuries, and it captures the main concerns of those who have favored state intervention in the education market.

Improving Morality of the Poor

Public education is often claimed to improve the lot of the poor. Some economists have felt that government education can help the poor by improving their morals.[3] Adam Smith (1937, p. 735) advanced this argument. He thought that industrial society, through the division of labor, encouraged a "drowsy stupidity, which . . . seems to benumb the understanding of almost all the inferior ranks of people." To remedy this mental torpor, Smith (p. 737) thought government could "encourage, and can even impose upon almost the whole body of society, the necessity of acquiring the most essential parts of education."

Thomas Malthus (1872, pp. 437–40), Nassau Senior (see Blitz 1971, p. 43), and Alfred Marshall (1920, p. 180) expressed similar views. All thought that poor parents generally do not have the foresight and strength of character to adequately provide for their children's education. They recommended compulsory education and government finance as a remedy.

A notable argument for state education along these lines was advanced by John Stuart Mill. Although Mill generally favored free choice and market competition, he thought that the consumer could not competently judge education. Wrote Mill (1973, pp. 953–54), "The uncultivated cannot be competent judges of education. Education, therefore, is one of those things which it is admissible in principle that a government should provide for the people."

---

[3] "Government education" here does not mean simply the direct government supply of education. Rather, it means any government involvement in the financing, consumption, or provision of education.

This idea finds modern expression in the work of John Vaizey. According to Vaizey (1962, pp. 28, 32), "Many studies have shown that parents make choices that are ill-informed and not in the children's own interests," and state education "can in a substantial degree be regarded as an intervention to save the individual from the family." (See also Benson 1968, p. 161.)

Despite its long standing among economists, the argument that state education can morally improve the poor has several weaknesses. A patronizing attitude runs through the arguments. The poor are variously described as stupid, profligate, weak, and uncultured. The wealthier and governing classes are assumed to possess none of these faults. On the contrary the virtues of the rich qualify them as models and teachers of the poor.

If we add paternalism to this patronizing attitude, state education follows as a matter of course. Forcing the poor to learn values that they would not learn voluntarily seems a mark of beneficence. To someone who does not see the rich and powerful as morally superior to the poor, however, compulsion appears otherwise. State education is a means by which the upper classes control the lower ones. The use of public education to exercise social control over the poor has been one of the main themes of American revisionist historians (see, for example, Katz 1971).

Depriving people of their right to choice because they are incompetent is a practice that can be applied to almost any activity, including religion. Herbert Spencer (1888, p. 367) pointed this out: "A stock argument for the state-teaching of religion has been that the masses cannot distinguish false religion from true." He also thought it "strange that so judicious a writer [as Mill] should feel satisfied with such a worn-out excuse. This alleged incompetency on the part of the people has been the reason assigned for all state-interferences whatever."[4]

The authoritarianism of instilling values in the poor through state education has been pointed out by E. G. West (1964). Government financing of education in Britain grew out of Jeremy Bentham's fondness for controlling the environment of the child. The intention was benevolent, says West, but the reality was authoritarian.

---

[4] The elder Mill had, several years before his son declared the uncultured to be unfit for judging education, ridiculed the English clergy for claiming that laymen were unfit judges of religion and therefore ought not to be permitted to read the Bible. James Mill pointed to the beneficial effects of allowing each man freedom to judge religious issues for himself. It seems ironic that John Stuart Mill, who was such a strong defender of religious freedom, would oppose educational freedom. (See Lively and Reese 1978).

West (1965, p. 211) also points to an inconsistency in the incompetent consumer argument. If members of a democracy are incompetent to judge education, how can they be competent to judge the representatives who will determine the education?

Instilling Values Essential to a Good Society

Besides arguing that it would benefit the poor, economists have maintained that state education should instill values that are essential to the existence of a good society. This argument, too, was advanced by Smith. In addition to believing that the division of labor induces a drowsy stupidity in workers (as noted above), he also believed that repeating routine tasks fosters cowardice. Stupidity combined with cowardice means that workers, who constitute most of the population, are unwilling and unable to defend their country in time of war. To prevent the intellectual and moral decay of industrial society, Smith (1937, pp. 736–40) thought the government should pass compulsory education laws and help finance education. This argument was later endorsed by both Malthus and Say (Malthus 1872, p. 441; Say 1964, p. 435).

Economists have been particularly concerned that state education teach values necessary for the smooth functioning of democracy. John Stuart Mill thought workers should be denied the vote unless they have a minimum of education (see Garforth 1980, p. 120). More recently, Charles Rowley and Alan Peacock (1975, pp. 127–28) have argued that state finance through vouchers is required to teach the common values necessary for democracy. Voluntary education, they believe, would result in values too diverse for democracy to work. Charles Benson (1968, p. 35) lists schooling's contribution to democracy as one of the "important, possibly overriding, benefits of school services" and G. U. Papi (1966, p. 8) writes, "If private enterprise alone were responsible for this service [that is, education] society would run the risk of receiving a kind of education not in line with national interests." Garms, Guthrie, and Pierce (1978, pp. 47–49) express similar ideas.

The arguments that state education should promote values essential to a good society can be interpreted two ways. First, without government education, proper values cannot develop and spread. Therefore, a good society is impossible without government education. Second, purely voluntary education could create and teach values that foster a good society, but government support of education can improve on voluntary means. These two interpretations are both plausible readings of the arguments, but each has sufficiently different implications that it is useful to consider them separately.

If it is accepted that voluntary teaching of values cannot produce a good society but that government teaching can, government should have ultimate authority over the dissemination of ideas. Government should not only encourage proper values, it should also discourage wrong ones. If voluntary education produces values too diverse for the functioning of good society, or if it teaches values so inimical to the national interests that a good society is not possible, the state should use its coercive power to impose uniformity and oppose heresy (see Spencer 1888, p. 365).

To those who value a free society, this argument has ominous implications. The authority of the state to oppose the voluntary exchange of ideas compromises freedom of religion. Henry Sidgwick (1891, pp. 557–60), who clearly saw this conflict between church and state, thought that government should have the power to control religion. Government would not have to exercise this power as long as church teaching did not conflict with government. If these values were to conflict, though, it would then be proper for government to suppress religious teaching.

The same remarks apply to the press. If newspapers and periodicals produce values too diverse for society to function, and government has the responsibility to see that values are sufficiently uniform, government must have the power to impose uniformity.

If voluntary education is able by itself to create a good society, the power of the state need not be so great. The state can limit itself to the subsidiary role of fostering the voluntary exchange of values. In this case, however, state education runs the danger of obstructing an educational system that if left alone, could produce a good society. As a subsidiary institution, state education may harm as well as help.[5]

Economists have expressed a great deal of concern about the harm that government can inflict on private education. Adam Smith opposed complete government financing of education, because it would dilute the incentive for good teaching. The greatest danger of state education, however, is its ability to establish tyranny.

J. B. Say expressed concern about this danger. In his writing (1964, p. 434), he pointed to the French universities under Napoleon that taught "opinions calculated to perpetuate the political slavery of their country." Mill sounded the most serious alarm about the potential for despotism. "A general State education," he wrote (1956, p. 129),

---

[5] See West (1970, p. 89).

"is a mere contrivance for moulding people to be exactly like one another . . . . in proportion as it is efficient and successful, it establishes a despotism over the mind, leading by natural tendency to one over the body."

If voluntary schooling alone can institute the values required of a good society, government education should not be a dominant force in education. The inefficiency of public schools, and especially the danger of government dominance over values, precludes consideration of any major role for state education. Of course, this still leaves room for state education to play a modest role. A voucher system, it may be argued, could be effective in helping to produce common values without running any great danger of gross inefficiency or government dominance.

Even this is doubtful, though, for modest government intervention in voluntary education leads toward government dominance. It would be naive to think that the state would not use such power as it had to secure more power. State teachers will preach the virtues of state education and the vices of private education. Public educators will extol the achievements of public education in self-serving histories (see Bailyn 1960, pp. 8–9). State school officials will try to expand the influence of public schools and restrict that of private schools. The attempt to outlaw private schools in Oregon is a striking example of this kind of restriction (see *Pierce v. the Society of the Sisters* in Reutter and Hamilton 1970, p. 501). The enormous growth of American public schools since their modest beginnings demonstrates how successfully government education can crowd out private education. A system that started with the modest financing of education has grown into an enormous system that compels attendance, licenses teachers, dictates curriculum, and dominates supply.

Even if a large degree of state control over education is never achieved, government schools secure their resources by force. If we endorse public schools, or public finance, or compulsory school-attendance laws, we also endorse the use of coercion to disseminate social values. Without state education, competition of ideas uses resources secured solely by consent. Also, if moderate amounts of state schooling improve social values, why not moderate amounts of state activity in other mediums that spread ideas? If we should have state-supported schools, we should also have state supported newspapers and churches.

Doubt is also cast on the value of moderate government schooling if we consider what "common values" it is supposed to promote (Rowley and Peacock 1975), or what it is supposed to contribute toward "political democracy" (Benson 1968) or toward the "national interest" (Papi 1966).

In modern liberal societies, people hold widely divergent views on religion, morals, politics, esthetics, and many other subjects. A liberal society permits, even encourages, diverse opinions and values. Common values are reduced to a minimum, and they consist of a respect for other people and their property. It is difficult to see how a system that coerces people and takes property by threat of force could improve on a system that honors property and requires the consent of the educated.

Furthermore, in both Britain and the United States, political democracy was established without public education and in the face of harsh opposition from governments. Agitation for freedom did not come from men educated in public schools. The Levellers, John Locke, Adam Smith, George Mason, and Thomas Jefferson were able to disseminate values that led to political democracy because their education was not in line with so-called national interests—that is, the interests of the established government.

Reducing Crime, Achieving Greater Social Equality

Two other ethical arguments favoring state education deserve brief treatment. One is that government involvement can reduce crime, the other is that it can reduce inequality.

Thomas Malthus was one of the early classical economists who advocated state education to reduce crime. Malthus felt it is the duty of government to provide for the instruction of its people. He thought it a disgrace that the education of the poor had been left to voluntary Sunday schools, and he believed that a proper education, financed by government, would reduce crime by improving morals. This idea finds modern expression in the Robbins report, which listed a "healthier and less crime-prone population" among the benefits of education (quoted in West 1965, p. 32). (See also Lott 1984, pp. 22–24, for other economists who have believed that public education reduces crime.)

Some economists, however, have doubted the ability of state schooling to improve morals or reduce crime. Say (1964, pp. 437–38) thought "The only important science not susceptible of being taught at the public charge, is that of moral philosophy." Say believed moral principles were best taught through voluntary social intercourse. Herbert Spencer (1888, p. 383) concurred: "Did much knowledge and piercing intelligence suffice to make men good, then Bacon should have been honest and Napoleon just."[6]

---

[6] For a discussion of Francis Bacon's character, see Buckle (1863, pp. 50–55).

More recently, some economists have suggested that public schooling may promote crime. Edwin West (1965, p. 36) has pointed out that when Britain raised the compulsory school-attendance age from 14 to 15, crime committed by 15-year olds increased markedly. Although West claims only that these statistics cast doubt on the assertion that education reduces crime, they can easily be interpreted as showing a link between increased government education and increased crime. In a recent doctoral dissertation, John Lott (1984, pp. 115–36) has found a strong correlation between changes in the percentage of students in public schools and changes in the juvenile crime rate. Lott's study attributes this to the better instruction of private schools. Instruction that is more valuable to students makes it more costly to commit crimes for which the penalty is expulsion from school.

Reducing social inequality has been given as a reason for state education by Henry Sidgwick, James Meade, and Charles Benson. Sidgwick (1891, pp. 152–53) thought public expenditure would transfer wealth from the richer to the poorer; on the assumption that a given sum of money gives more utility to a poor man than to a rich one, this transfer would increase social utility. For utilitarians, a maximum of social utility was a primary goal of public policy. Meade's reasoning is similar to Sidgwick's, except that he does not believe social utility can be measured. Meade (1976, pp. 195–99), thinks the heavier tax burden on the wealthy is an efficient way to transfer wealth.

Social equality finds its most important justification in the work of Benson, for whom social equality is the main justification for the government to supply education. Government-operated schools increase equality by increasing social mobility. According to Benson (1968, p. 59), "The justification for public *operation* of the schools rests, then, on the contribution of the public school system to preserving social mobility."

This objective will not be attained if the government runs schools only for the poor. If the poor were to attend public schools, and everyone else went to private schools, public schools would fall into disrepute. However, according to Benson (1968, p. 60), "With predominantly *one type* of education in the United States, no student need feel he is handicapped with respect to advancement in income and status by the fact that he is attending, (or has attended) the type of school called public. . . . This, we believe, is the primary case for public operation."

The arguments for social equality have several weaknesses. For Sidgwick and Meade, social equality means equality of income. Sidgwick advocated equality because a given sum of money transferred from a wealthy man to a poor man increased the utility of the poor man by more than it decreased the utility of the wealthy one. Since Sidgwick's time, economists have abandoned their belief in measurable utility and consequently, have given up making interpersonal utility comparisons.

Meade does not attempt to justify income equality. He simply assumes it is desirable and focuses on the most efficient way to achieve it. Because Meade does not present an argument, it is not useful to criticize the pursuit of income equality by coercive transfer. (On the impracticability and injustice of income equality, see Nozick 1974, pp. 160–64.)

Social equality to Benson means equality of education. By having government operate all the schools, he wants to eliminate differences in income that are attributable to differences in the quality of schooling. Equality of education, however, does not logically entail government-run schools. The same result could be achieved by imposing a uniform curriculum in all schools, by compelling everyone to complete the curriculum, and by subsidizing those who could not afford to pay. Any school that offered better courses than others, or that taught the standard curriculum better than others, could be closed down. This would achieve the same educational equality through regulation that Benson would achieve by government operation.

Government-administered schools are not only unnecessary to achieve educational equality but also are insufficient. Some government schools are better than others. Schools in wealthier suburban areas are usually better, and have better reputations, then inner-city schools. To achieve educational equality, all government schools would have to conform to uniform standard, too.[7]

The drawback to this is obvious. By requiring all schools, whether government or private, to be uniform, by removing the incentive for one school to improve faster than others, by penalizing those people who are able and willing to teach better than others, equality of education would prevent the most capable educators from exercising their talents. Achieving Benson's equality would subject every student in the United States to the worst schools and teaching available, and it would limit the pace of improvement to that of the slowest educators in the system.

Eugenia Toma has recently pointed out yet another drawback to uniformity. By requiring all schools to be uniform, government reduces competition among producers to supply the kinds of education consumers want. The result has been a greater monopoly control of education, with higher salaries and benefits accruing to public educators. The rhetoric for equality of educational opportunity has translated into monetary gains for public educators.[8]

---

[7] See Friedman and Friedman (1980, p. 169); and West (1985, pp. 31–32).

Increasing Economic Production

The modern argument justifying public education as a means to increase economic productivity was first drawn by Sidgwick and Arthur C. Pigou. Sidgwick (1891, pp. 147–48) argued that the higher productivity afforded by more education could not be exploited by poor children, because their parents could not afford to pay for the education and could not secure loans, even though the increased productivity would more than repay such loans. Pigou added to the argument by noting that educational loans for poor children were difficult to obtain because capitalists, unable to own what they were investing in, could not adequately secure their loans. "Under a slave economy," observed Pigou (1920, p. 776), ". . . the case would be different." However, by taxing the rich and making the money available to the poor to invest in themselves, Pigou believed that output would be significantly expanded. Underinvestment in education owing to "transactions costs" has now become known as the "imperfect capital market" argument, and it is widely used to justify government loans to students.[9]

It is doubtful that a policy of taxing the rich and using the tax receipts for educating the poor would increase productivity. As Adam Smith pointed out, such a policy will decrease the incentive of the schools to educate efficiently (see Smith 1937, pp. 717–21). This inefficiency is increased when tax money supports a large and politically powerful state educational system. The government must observe how and to what extent it alters incentives if it wishes to employ capital more efficiently than the private market.

Loans made directly to students will probably have the least effect on the incentives of the schools. Direct loans from government, however, also alter incentives. The risk of loaning against increased future earnings is only partially attributable to the prohibition of slavery. Identifying those who will succeed is also risky. This risk is compounded by the difficulty of knowing which capital goods to tax. It is not enough to pass a general tax on property or capital goods. Capital must be diverted from particular investments that have a lower return and invested in particular people who will yield a higher return. Such diversion and investment requires a high degree of skill and judgment.[10] The market has a selection process that places these decisions in the hands of the most competent. This is perhaps the most important source of efficiency in a market.

---

[8] See Toma (1983).

[9] See Friedman (1955, pp. 139–42); Arrow (1968, p. 878); and Miller (1967, p. 200).

Government officials are not tested by the competition of the market. Rising to positions of influence in the bureaucracy does not test a person's ability to successfully invest scarce capital. In competent hands, it may be possible to increase output, as Pigou (1920, p. 782) believed, but there is no reason to expect competent hands to capture the post.

No matter who succeeds in deciding who to tax or where to invest, public officials do not have the same financial incentive to be efficient that private investors have. Neither their careers nor their financial success depends on taxing and investing efficiently. The present system of financing state-run schools through local property taxes, as well as general state and federal tax funds, demonstrates how little incentive government educators have to tax efficiently. These taxes make no pretense of taking money only from less-productive investment projects.

The diminished incentive of public officials to be efficient also extends to the collection of loans. Less government incentive to collect loans increases the default rate. If the increased rate of default is not offset by a sufficiently high return on loans that are repaid, government financing will decrease output.

Finally, we draw attention to the moral implications of justifying government financing by imperfect capital markets. Pigou (1920, p. 776) noted that the absence of slavery in a market economy increases the risk of investing in education. If the rich could indenture the poor, or own them outright, risk would be reduced and output increased. Enslavement of the poor by the rich is repugnant because we value freedom above increased productivity. This same value scale would also prohibit the coercive transfer of resources to merely increase output. Combined with the difficulty of judging where to diminish investment and where to increase it, and with the attenuated incentive to tax and invest efficiently, the moral implications of using force to increase production militate against government finance of education to increase output.

---

[10] A knowledge of average returns does not indicate whether the government transfer increases or decreases efficiency. An average return of 10 percent on capital goods compared with an average return of 12 percent on educational investment can result from taxing capital goods projects that would have earned 14 percent. A knowledge of marginal returns—that is, returns on particular projects—is required to compare efficiencies.

Internalizing External Benefits

The existence of external benefits is the main reason that modern economists favor government education. Like the other arguments examined here, though, this one has roots that extend back to Adam Smith and the classical economists.

Smith (1937, p. 768), in his praise of the Scottish school system, wrote: "The expense of the institutions for education and religious instruction, is likewise, no doubt, beneficial to the whole society, and may, therefore, without injustice, be defrayed by the general contribution of the whole society." However, Smith did not insist that everyone pay for education. His next sentence (p. 768) reads: "This expense, however, might, perhaps with equal propriety, and even with some advantage, be defrayed altogether by those who receive the immediate benefit of such education and instruction, or by the voluntary contribution of those who think they have occasion for either one or the other."

J. B. Say was more insistent on tax support of some kinds of education. According to Say (1964, p. 433), "[I]t is not every degree or class of knowledge, that yields a benefit to the individual, equivalent to that accruing to the public." Those who study the general laws of nature and who write good elementary treatises are not paid in proportion to the benefit they confer. Therefore, thought Say (pp. 328–29, 433–35), government should support schools and recompense authors of good textbooks.

John Stuart Mill (1973, p. 958) concurred that scientific research benefits the whole community and is therefore a service "for which it is *prima facie* reasonable that the community should collectively pay," but it was Henry Sidgwick who added substance to the argument. If scientific research benefits the general populace, asked Sidgwick (1891, p. 142), why do the people not voluntarily pay for it? His answer was that transaction costs are too high for voluntary exchange to occur. Government expenditure can circumvent these costs, thereby increasing general welfare.

Sidgwick also thought that educating the poor benefited the entire community by generally improving morals and intellect. He urged the state to finance both religious and secular teaching.

Milton Friedman (1955) has been the most influential modern economist to advocate government financing of education to secure external benefits. Friedman's argument runs as follows. An educated person possesses literacy and knowledge that make possible a stable, democratic society. We all benefit from this kind of society, but we do not pay the educated person for this benefit. The proper financial incentives to acquire education are therefore lacking, and too little education will be acquired.

For Friedman (1955, pp. 124–25), the external benefits of education justify laws that "require each child to receive a minimum amount of education of a specified kind." Government financing is also justified as "the only feasible means of enforcing the required minimum." (Friedman, however, later changed his mind about this after considering Edwin G. West's arguments against state education. (See West 1976, pp. 92–93.)

This general argument is widely accepted as economic justification for state-financed education. Benson (1968), Wiseman (1959), Rowley and Peacock (1975), Papi (1966), Garms (1978), Weisbrod (1964), and Cohn (1979) are among the modern economists who accept this argument.

The existence of external benefits does not justify the compulsory financing of education. As Edwin G. West (1965, p. 266) was first to point out, even though we may benefit from another person's education up to a point, we may not gain anything from the additional education that would be acquired through state financing. If people voluntarily acquire enough education to produce a stable and democratic society, any additional education that results from state financing will not benefit us in this way. Thus, according to West, even though the total external benefits of education are substantial, the marginal external benefits can be negligible.

Moreover, even if we were to benefit from the additional schooling of another person—that is, even if there were marginal external benefits—government financing would not necessarily be justified. Although we may benefit from the additional education of others, we may not benefit enough that we want to pay for it. If we have more important uses for our income, the external benefits are not Pareto relevant, and government financing is not justified.[11]

Most of the things we buy have marginal benefits, but we do not buy additional units (another car, for example) because we have more pressing uses for our income. By forcing us to pay for other people's education when we have more important uses for our income, government reduces rather than enhances the general welfare.

---

[11] For a classification of externalities, see Buchanan and Stubblebine (1962, pp. 371–84).

Granting that the marginal external benefits to education are substantial and Pareto relevant still does not justify government financing, because we can pay for those benefits voluntarily. The establishment of funds for research, scholarship, and teaching are examples of one person paying for the education of others to capture external benefits. If we do in fact benefit from the additional education of others, voluntary contributions enable us to capture these benefits without the necessity of government action. To justify government financing, we must explain why we do not pay for the education of others when it is worth the price to us.

The answer given by Henry Sidgwick was that transaction costs are too high for voluntary exchange. Government can increase welfare by taxing us and by paying for those things that we would pay for voluntarily if only the costs of making the exchange were not prohibitive.

This argument implicitly assumes that government officials do not have high costs of transacting at the same time that private persons do. It also assumes that government officials want to take from us in taxes only what we would voluntarily spend on education were transactions costs not so high. Neither of these assumptions is plausible.

This discussion of Pareto-relevant externalities illustrates the high costs of coercive transfer that aims at increasing the general welfare. To increase welfare through financing education, the government must determine how much each person would voluntarily spend on education were it not for high transactions costs. The government would have to avoid taxing those for whom marginal external benefits were not Pareto relevant, and it would have to tax all others in accordance with the value of the marginal external benefits. It would then have to enact laws, collect taxes, and dispense the funds properly. Economists who have argued that external benefits justify the government financing of education have not addressed the enormous information costs that would confront such a system. They have not explained in any detail how governments would assess, levy, and disperse taxes so as to identify and remove Pareto-relevant external effects. They have not even specified in any detail the extent of Pareto-relevant externalities. In consequence the existence of Pareto-relevant externalities is little more than assertion.

This discussion of transaction costs has assumed that government officials try to increase welfare by levying taxes according to the value of external benefits. This assumption, however, ignores the economic incentives created by taxation. Once educational revenue is obtained through taxes, those whose salaries and positions depend on educational expenditures have a financial incentive to increase taxes, regardless of whether those who are taxed benefit. Also, parents and students can gain financially by forcing others to defray a part of their educational expenses. Government financing of education creates a group whose interest lies, not in enhancing the general welfare, but in reducing it.[12]

The external benefits argument is also inconsistent with freedom of speech and religion. Books, newspapers, television, and radio contribute as much as—and perhaps more than—school to a stable and democratic society. If government should subsidize schools, it should by the same logic subsidize the press. The opportunities for a government-financed press to aggrandize the state are widely recognized, and freedom of the press is jealously guarded. Government-financed schools present similar opportunities for state propaganda, but the dangers are less appreciated.[13]

The external benefits argument of education can be applied directly to religion. Christian values, by condemning violence and teaching respect for others, help create a stable and peaceful society. These values benefit not only the person who holds them but his neighbors as well. Therefore, we could argue, the state should compel everyone to learn Christian values, and it should subsidize religion. If compulsory education and government financing of schools are justified by external effects, so are compulsory religion and subsidized churches. With the exception of Sidgwick (1891, p. 148), economists have not been consistent enough to apply their educational arguments to religion.

Summary and Conclusion

In the 200-year sweep of economic thought, economists have supported some state involvement in education, but have also been chary of the dangers of such involvement. Their concerns with morals and efficiency have led economists to recommend that the state compel, finance, and supply education, but these same concerns have also led them to advise against compulsion, finance, and supply.

The preponderant weight of economic opinion probably favors state finance and compulsion, but goes against direct government supply. This paper has tried to show that the arguments favoring any state activity in education are not compelling. It has argued that neither morals nor efficiency are likely to be improved by introducing compulsion into the education market.

---

[12] In general, government intrusion into the market will create as well as eliminate external effects. Government is a method of obtaining benefits that the recipient does not have to pay for. The costs of obtaining the benefits are therefore "external" to the recipient (see Kalt 1981, pp. 578–79.)

[13] For an example of state educators attempting to squelch the teachings of economists to protect special privileges, see Frederic Bastiat's "Declaration of War Against the Professors of Political Economy," in Bastiat (1964, pp. 294–300).

The arguments presented here against state involvement can be separated into three main strands. First, it has been pointed out that governments have no special claim to enhancing social efficiency. They have no special expertise in solving the informational problems that create social inefficiency. Second, there are the perverse incentives created by government involvement. Government officials do not have private interests that correspond to those of the public. Their incentives will in fact lead to social inefficiency. Third, it has been shown that government involvement in education endangers the traditional values of a free society. The same logic that urges state education on society also encourages state religion and a state-run press.

Of course, there may be compelling arguments for state education other than those advanced by economists. Economists though have set forth a great many arguments, both moral and economic. By demonstrating that their arguments are neither logically compelling nor consistent with traditional American freedoms, this paper has gone some distance in showing that education is best left to the voluntary choices of the market.

References

Arrow, Kenneth. "Criteria for Social Investment." In *Readings in the Economics of Education*. Paris: UNESCO, 1968.

Bailyn, Bernard. *Education in the Forming of American Society*. New York: Vintage Books, 1960.

Bastiat, Frederic. *Selected Essays on Political Economy*. Princeton: D. Van Nostrand Co., 1964.

Benson, Charles. *The Economics of Public Education*. 2d ed. Boston: Houghton Mifflin Co., 1968.

Blitz, Rudolph C. "Education in the Writings of Malthus, Senior, McCulloch and John Stuart Mill." In *Readings in the Economics of Education*. Paris: UNESCO, 1971.

Buckle, Henry T. "Mill on Liberty." In *Essays*. New York: Appleton, 1863.

Buchanan, James, and Stubblebine, William Craig. "Externality." *Economica*, n.s. 29 (November 1962): 371–84.

Cohn, Elchanan. *The Economics of Education*. Cambridge, Mass.: Ballinger Publishing Co., 1979.

Friedman, Milton. "The Role of Government in Education." In *Economics and the Public Interest*. Edited by Robert A. Solo. New Brunswick, N.J.: Rutgers University Press, 1955.

Friedman, Milton, and Friedman, Rose. *Free to Choose.* New York: Harcourt Brace Jovanovich, 1980.

Garforth, Francis W. *Educative Democracy.* Oxford: Oxford University Press, 1980.

Garms, Walter; Guthrie, James; and Pierce, Walter. *The Economics and Politics of Public Education.* Englewood Cliffs, N.J.: Prentice Hall, 1978.

Johnson, E. A. J. "The Place of Learning, Science, Vocational Training, and 'Art' in Pre-Smithian Economic Thought." *Journal of Economic History* 24 (June 1964): 129–44.

Kalt, Joseph. "Public Goods and the Theory of Government." *Cato Journal* 1 (Fall 1981): 565–84.

Katz, Michael B. *The Irony of Early School Reform.* Boston: Beacon Press, 1968.

Katz, Michael B. *Class, Bureaucracy, and Schools.* New York: Praeger Publishers, 1971.

Lively, Jack, and Reese, John. *Utilitarian Logic and Politics.* Oxford: Oxford University Press, 1978.

Lott, John. "Alternative Explanations for Public Provision of Education." Unpublished doctoral dissertation, University of California at Los Angeles, 1984.

Malthus, Thomas R. *An Essay on the Principle of Population.* 7th ed. London: Reeves and Turner, 1872.

Marshall, Alfred. *Principles of Economics.* 8th ed. London: Macmillan and Co., 1920.

Meade, James. *The Just Economy.* London: George Allen and Unwin, 1976.

Mill, John Stuart. *On Liberty.* New York: The Bobbs-Merrill Co., 1956.

Mill, John Stuart. *Principles of Political Economy.* Clifton, N.J.: Augustus M. Kelley, 1973.

Miller, William. "Education as a Source of Economic Growth." *Journal of Economic Issues* 1 (December 1967): 280–96.

Nozick, Robert. *Anarchy, State and Utopia.* New York: Basic Books, 1974.

Papi, G. U. "General Problems of the Economics of Education." In *The Economics of Education.* Edited by E. A. G. Robinson and John Vaizey. London: Macmillan and Co., 1966.

Petty, William. *The Economic Writings of Sir William Petty.* Edited by Charles Hull. Cambridge: Cambridge University Press, 1899.

Pigou, Arthur C. *Economics of Welfare.* London: Macmillan and Co., 1920.

Reutter, E. Edmund, and Hamilton, Robert R. *The Law of Public Education.* Mineola, N.Y.: The Foundation Press, 1970.

Rothbard, Murray. "Total Reform: Nothing Less." In *Nonpublic School Aid.* Edited by Edwin G. West. Lexington, Mass.: D.C. Heath and Co., 1976.

Rowley, Charles, and Peacock, Alan. *Welfare Economics.* New York: John Wiley and Sons, 1975.

Say, Jean Baptiste. *A Treatise on Political Economy.* New York: Augustus M. Kelley, 1964.

Sidgwick, Henry. *The Elements of Politics.* London: Macmillan and Co., 1891.

Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations.* Modern Library ed. New York: Random House, 1937.

Smith, George. "The Voluntaryists: Prophets of Modern Revisionism." In *The Public School Monopoly*. Edited by Robert B. Everhart. Cambridge, Mass.: Ballinger Publishing Co., 1981.

Spencer, Herbert. *Social Statics*. New York: D. Appleton and Co., 1888.

Steuart, James. *Collected Works*. London: T. Cadell and W. Davies, 1805.

Toma, Eugenia. "Institutional Structures, Regulation, and Producer Gains in the Educational Industry." *Journal of Law and Economics* 26 (April 1983): 103–16.

Vaizey, John. *The Economics of Education*. London: Farber and Farber, 1962.

Weisbrod, Burton. *External Benefits of Public Education*. Princeton: Princeton University Press, 1964.

West, Edwin G. "Private vs. Public Education: A Classical Economic Dispute." *Journal of Political Economy* 72 (October 1964): 465–75.

West, Edwin G. *Education and the State*. London: Institute of Economic Affairs, 1965.

West, Edwin G. "Resources Allocation and Growth in Early Nineteenth-Century British Education." *Economic History Review*, 23 (April 1970): 68–95.

West, Edwin G. *Nonpublic School Aid*. Lexington, Mass.: D.C. Heath and Co., 1976.

West, Edwin G. "The Demise of 'Free' Education." *Challenge* (January–February 1985): 31–32.

Wiseman, Jack. "The Economics of Education." *Scottish Journal of Political Economy* 6 (February 1959): 48–58.

PRIVATIZATION: A STRATEGY TO CUT THE BUDGET

Stuart Butler

Introduction

Even the most die-hard supporters of Ronald Reagan are waking up to a very unpleasant fact: He has not cut the federal budget—in real or nominal terms—and there is little reason to suppose that he will do so in the remaining years of his tenure at the White House. As a proportion of GNP, spending surged in the first three years of Reagan's initial term, ending any hope of even bringing spending down to the proportion of output consumed by Washington during the Carter "big spending" years, let alone achieving Reagan's original objective of a dramatic rollback of the federal sector. And even though the step-up in military outlays was clearly an element in the growth of government, that was factored into the original White House spending projections. The real problem has been a failure to win congressional approval for significant cuts in domestic social programs.

Some social programs have certainly been cut. To judge from the media coverage, one might suppose that these cuts have been draconian. Far from it. Means-tested welfare spending actually increased over 7 percent between 1980 and 1984. More important, the entitlement programs benefiting middle-class Americans, including Social Security, Medicare, and federal retirement, expanded by more than 17 percent in real terms during Reagan's first term.

This inability to turn the spending tide slowly transformed David Stockman from a can-do *wunderkind* into a frustrated and embittered cynic. Those who still urged a determined attack on entitlement

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Director of Domestic Policy Studies at The Heritage Foundation. This paper is based on the author's recently published book *Privatizing Federal Spending* (New York: Universe Books, 1985).

programs he dismissed as "dreamers" in a 1984 *Fortune* interview. Forget the idea of getting spending down below 20 percent of GNP, said the budget director, "The minimum size of government achievable appears to be 22 percent to 23 percent of GNP."[1]

The easiest reaction to this utter failure to cut spending would be to join the chorus now excoriating the White House and its congressional supporters for their supposed lack of resolution in holding the budget line. Yet it is hard to fault the Reagan administration on its determination to follow through with the 1980 election pledge to reduce the size of government. The lost Battle of the Budget has had nothing to do with any absence of will—but it has much to do with obsolete political weaponry. The critical strategic mistake of the administration was its assumption that the only way to contain and then cut the budget was to seek legislation to cut the supply of dollars leaving Washington—one might call it a "supply side" view of the budget. This led the White House to view spending in aggregate terms and to regard the budget problem solely as one of managing and restraining expenditures. Congressional tactics then centered on identifying program cuts that would add up to the aggregate target.

The fatal flaw in this supply-side view of the budget process is that it largely ignores the demand side of the political process. It overlooks, in other words, the subtle process by which government programs grow and are defended by the efforts of powerful coalitions. These coalitions do not compete for a fixed amount of spending set by Congress. Quite the opposite. Total federal spending is the cumulative result of successful political demands by these coalitions. Trying to cut down the size of government by holding down total outlays, without first tackling this constituency pressure, is like trying to stop a pot from boiling by clamping down on the lid.

The Reagan administration will prove no more successful in its second term at cutting spending unless it devises a strategy to block or divert the demand pressure for spending. Only if the constituency heat is turned down will congressional supporters of smaller government be able to vote their principles—and be reelected. To change the underlying political dynamics of spending, the administration should look hard at the enormous potential of privatization, that is, the shifting of government functions to the private sector. So far the Reagan administration has viewed privatization simply as a way of bringing down the cost of government programs by utilizing more efficient private-sector suppliers. But a closer look at the dynamics of the budget reveals that privatization could become a potent political weapon to eliminate programs. With suitable incentives as a catalyst, a privatization strategy would create "mirror image" coalitions in opposition to the constituencies now pressing for federal spending, which would divert the demand for many services out of the public sector and into the private sector. Privatization thus offers an opportunity to turn down the political heat for federal programs.

---

[1] "David Stockman: No More Big Budget Cuts," *Fortune*, 6 February 1984, p. 54.

Why Spending Increases: The Federal Ratchet

To appreciate why privatization could be so effective, it is first important to understand how spending increases and why programs are so resistant to demands for cutbacks.

Government programs tend to grow according to a now-familiar pattern. The first step is the creation of a relatively small program to aid a limited group. This might be a group of distressed or disadvantaged Americans, such as impoverished single parents with dependents (note more recently, pressure to provide federal aid to the homeless). It might involve a specific industry or sector with supposedly special problems, such as agriculture or small business. Or it might be a traditional All-American pork-barrel project to help politically sensitive areas of the country, such as urban development assistance. Normally the budget allocation is so small that the program can be financed without any significant extra burden on the individual taxpayer.

Once this initial funding has been voted and the new program created, however, a constituency soon coalesces around the program. The resulting coalition consists of three powerful elements:

1. *Beneficiaries and "Near" Beneficiaries.* New programs rarely satisfy the expectations of the targeted beneficiaries: The original Social Security benefits seemed meager after a while, for instance, and recipients of food stamps never have quite enough to eat. So it is not long before beneficiaries of programs form organizations and enlist lobbyists to press for more spending. Moreover, there are inevitably those who do not quite make the eligibility criteria of the original program, yet still feel they meet the intent of the law: They may have just one more employee than the criterion set for "small" business assistance, or be a poor neighborhood in an otherwise prosperous city ineligible for development grants. These "near" beneficiaries, as one might call them, soon begin to argue vociforously that they have been "unfairly" excluded from the program. Usually it is far less painful, politically, for Congress to expand the program to include them than to spread existing funds more thinly. But once

this group of near beneficiaries is accommodated, another group invariably appears.

2. *Administrators*. More federal spending means more jobs and promotion opportunities for federal employees who administer programs. So officials have good reasons to join with beneficiaries, albeit discreetly, in lobbying for the expansion of the programs that employ them. And even if officials are not motivated by self-interest, the very fact that they are in government service tends to lead them to the conclusion that government—indeed their own agency—must play a significant role in solving the country's problems. As Martha Derthick notes in her study of the growth of Social Security: "The prevailing technique of policy analysis [within the Social Security Administration] was to identify a social problem, such as lack of health care, and to develop the arguments and methods for dealing with it through social insurance."[2]

3. *Service Providers*. The private sector is not immune from the lure of federal largesse. When roads are built with federal money, private construction firms prosper. When there is a defense buildup, that means jobs in the aircraft industry. When welfare spending is increased, a corps of human-service professionals can look forward to a more secure future. So segments of the private sector are inevitably drawn into the coalition lobbying for more spending.

These three-pronged coalitions provide the underlying momentum for increased federal outlays. And the balance of political advantage is with the spending coalitions. The outlays are concentrated on only a small segment of the entire population, yet the tax and borrowing cost of the outlays is spread over all taxpayers. Since most of these costs are not even in the form of explicit individual taxes, but are hidden in corporate taxes (passed on in higher prices), government borrowing, or inflation, the direct tax burden of any particular program on any particular taxpayer is usually trifling. So the taxpayer has little financial incentive to spend time or money opposing any one program, even if he grumbles about spending levels in general. Yet the coalition member often has every reason to engage in time-consuming and expensive campaigns to convince the public that preserving his program is essential to national well-being, or to win congressional approval for new outlays.

Success in this process is made easier by two techniques routinely used by the public-spending coalitions. First, whenever possible lobbyists press for a federal program rather than a state or local government program. The reason for this is very simple—the tax costs of benefits are spread wider and thinner in a federal program, and so the determination of any taxpayer to oppose the program is reduced. Second, sophisticated lobbyists make sure that they do not ask for everything at once. An expensive new program may be voted down in a head-to-head confrontation in Congress, but a modest program may seem insignificant. Yet it provides the nucleus around which a coalition can form, leading to a build-up of political pressure for incremental expansion.

---

[2] Martha Derthick, *Policymaking for Social Security* (Washington, D.C.: Brookings Institution, 1979), p. 25.

This balance of political forces, of course, only operates in the direction of increased spending. Cutting or, worse still, eliminating a program strikes hard at the interests of the concentrated coalitions supporting it. Yet savings achieved by cutbacks are spread thinly over all taxpayers—at least in theory. Usually they are not spread at all—it is a long time since Americans received a tax rebate because Congress cancelled a program. Consequently, when budget cuts are threatened, powerful lobbies spring into action. "Studies" are commissioned by the coalitions and their congressional supporters to show the calamitous results that would surely follow spending cuts, distraught witnesses are paraded through well-publicized hearings, and, for good measure, proponents of reductions are portrayed as anti-poor, dogmatic, soft-on-defense, or whatever epithet most effectively isolates them from good Americans. And as Ronald Reagan discovered when his budget request landed on Capitol Hill earlier this year, his backbone middle-class constituency becomes the most implacable defender of a federal role when its own programs, such as Social Security, are threatened.

The whole process is like a ratchet. Movement occurs only in one direction. The underlying political dynamics leads to "creeping" federal spending, as small new programs are quietly launched, additions grafted on to existing ones, and cuts blocked with fierce determination. Even when taxpayers are sufficiently aroused to demand cuts by Congress, supporters of an active federal government are careful to ensure that although outlays may be trimmed, as few programs as possible are actually eliminated. In this way the coalitions remain intact, ready to fight another day.

Constructing a Private-Sector Ratchet

Complaining about public-spending coalitions achieves little more than high blood pressure. But developing methods to entice the public to choose a private rather than a public way of promoting their self-interest may achieve a great deal. Privatization seeks to shift functions out of the government sector by using a carefully crafted set of inducements. The privatization strategy recognizes the existence of the federal ratchet and imbalance of political forces favoring spending. But instead of trying to win an unwinnable war of attrition over the budget, privatizers attempt to turn the political flank of the spending coalitions.

Privatization does not necessarily deny that government, as the instrument of society, has an obligation to ensure that goods and services should be provided for certain public purposes. But a distinction is drawn between government as a *provider* (implying that government should levy taxes and deliver services itself), and government as a *facilitator* (implying that it should encourage or require those services to be provided by the private sector). Privatization, in other words, means seeking to transfer programs into the private sector using the carrot of incentives, not the stick of aggregate cutbacks. Rather than trying to frustrate political demands for spending, it means *deflecting* that demand into the nongovernment sector.

In this emphasis on the consumer's reaction to incentives and the "price" of alternative ways of satisfying his demands, privatization is analogous to microeconomic analysis, whereas traditional budget management is more akin to macroeconomic attempts to manage the economy. Indeed, one of the leading privatization theorists, Madsen Pirie, coined the term "micropolitics" to distinguish the privatization approach from other budget strategies:

> Micropolitics is the art of generating circumstances in which individuals will be motivated to prefer and embrace the alternative of private supply, and in which people will make individual and voluntary decisions whose cumulative effect will be to bring about the desired state of affairs. The process of transfer from [the] public to [the] private economy is most securely achieved when its progress is evolutionary, arising from free decisions.[3]

While deflecting demand in this way constitutes the broad strategy, the political tactics to achieve it consist of using incentives and other devices to create coalitions of beneficiaries, near beneficiaries, service providers, and private sector administrators to lobby for an expanding private role in delivering what is now a federal program, just as the pro public-sector coalitions currently lobby for increased federal spending.

These privatization coalitions are the mirror image, so to speak, of the public-sector coalitions. And they are at the heart of the strategy to create a "privatization ratchet" to counter the federal ratchet. By providing a targeted benefit (such as a tax incentive or some regulatory relief) to those who demand or provide a private alternative to government, considerable rewards can be guaranteed to individuals within the coalition. Members of that coalition can be expected to press for deeper incentives and to oppose any move to eliminate existing incentives. Similarly, near beneficiaries can be counted upon to demand inclusion under the terms of the incentive. On the other hand, the cost of the incentives spread thinly and widely—even assuming, for the sake of argument, that tax relief for one group implies an additional burden for everyone else. So there is little will to resist the coalition's demands.

---

[3] Madsen Pirie, *Dismantling the State* (Dallas: The National Center for Policy Analysis, 1985), p. 29.

Privatization thus turns conventional political dynamics on its head. Lobbying pressure develops for less taxation (if a tax incentive is given), and for private, not public, programs. Moreover, each legislative victory won by the coalition, however small, serves to strengthen it, thereby adding to its capacity to achieve further legislative concessions and a corresponding growth in the private program.

As the privatization coalition grows, and constitutes an attractive alternative to government programs, the equivalent public-sector coalition becomes less powerful. Thus a government that assiduously cultivates privatization becomes increasingly able to confront proponents of government programs and to win budget cutbacks in the traditional way. Yet it is only by altering the underlying political balance, through a privatization strategy, that victories in Congress can be obtained.

The Case of IRAs

To appreciate how the politics of privatization work in practice, consider Individual Retirement Accounts (IRAs) and Social Security.

Social Security is a classic example of the federal ratchet in operation. Established in 1935 to provide Americans with modest retirement benefits financed by a payroll tax, Social Security has mushroomed into an enormous program accounting for one-fourth of all federal outlays, guaranteeing a retirement income, disability benefits, spousal benefits, and medical care. As Ronald Reagan has learned the hard way, any politician who even talks about trimming these benefits significantly invites the wrath of an extremely powerful constituency.

Yet a minor provision in the 1981 tax act may eventually break up that coalition, if the White House seizes the opportunity it has been given. By allowing all working Americans to open tax-deductible

IRAs, Congress planted the seeds of a private alternative to Social Security. The $2000 annual deduction now available is a concentrated benefit only available to those choosing to open these accounts, while the "cost" (again, assuming that there is a corresponding cost) is spread over all taxpayers, ensuring that the political advantage is with the supporters of IRAs.

It was not long after the passage of the 1981 act that banks and other financial institutions (the administrators and service providers of IRAs) began a massive campaign to encourage Americans to open retirement accounts. Soon after that, nonworking married women began to complain that limiting their deduction to just $250 was unfair and discriminatory (near beneficiaries). And politicians were quick to propose accommodating the near beneficiaries and increasing the standard IRA deduction. A privatization coalition was born. The "tax loss" (as the Treasury puts it) of IRAs has vastly exceeded the original Reagan Administration projections. Yet repealing or reducing the deduction is already politically unthinkable—the coalition is too powerful and the privatization ratchet is in place.

There can be little doubt that in the years to come, IRAs will be seen increasingly by Americans—especially younger workers—as a more attractive retirement plan than Social Security. The IRA industry has every reason to support this, while avoiding any direct criticism of Social Security. In short, the incentive has begun to divert the pressure of demand for a secure retirement income away from the publicly provided system (Social Security) and to the private alternative (IRAs). Any modest legislative concessions won by the IRA coalition (such as a supplemental "medicare" IRA or a full deduction for nonworking spouses) will enhance the attractiveness of the IRA and further deflect pressure. From the budget-cutter's point of view, the growing power of this IRA coalition offers the only real hope for spending reductions in Social Security, since the more Americans prefer IRAs as their primary pension vehicle, the weaker will become support for retaining Social Security in its present form.

Developing a Privatization Strategy

The techniques available to the privatizer are extensive and varied. Different combinations and emphases are appropriate for different types of programs. But there are six tactics that must form the foundation of a general strategy. The first three tactics aim at weakening the public-sector coalitions while the last three tactics attempt to foster private-sector coalitions.

1. *Press for Existing Programs to be Administered and Financed "Closer to the People."*

If it is impossible to shift a program out of government, every effort should be made to shift the funding and delivery of the program down the federal system as a precondition for privatization. It is no coincidence that privatization in America has been spreading most rapidly at the municipal level. When the costs of a service are imposed on a relatively small group of taxpayers, resistance to spending increases is stronger and the demand for economies is greater. The centralization of government services and funding sources has been critical to the success of public-spending lobbies. Decentralizing government is critical to successful privatization.

2. *Apply User Charges Whenever Possible.*

The more the user of a publicly provided service has to pay, the more attractive the private alternative becomes. When combined with a tax incentive favoring private provision, a user charge can be a strong inducement for the public to switch its demand.

3. *"Buy Out" Key Elements of the Spending Coalition.*

Detaching powerful elements of a coalition supporting public spending strengthens the relative position of the corresponding privatization coalition. Britain's Conservative government was able to thwart union opposition to the privatization of the nationalized telephone system last November—the world's largest stock offering—by giving employees a special discount price for shares. Ninety-five percent of the unionized workers eagerly bought shares, despite the union hierarchy's total opposition to the sale.

4. *Deregulate Government Monopolies.*

Shifting consumer demand into the private sector is not likely to be successful if private-service suppliers are barred from freely entering the market. Consequently deregulation of service supply is essential for privatization. It should be noted here that this requirement also applies when private-sector organizations supply services under contract to government. Arguments abound as to whether contracting out should really be called privatization (since the funding is governmental), but one thing is clear: private-sector suppliers immune from competition raise costs and erect barriers to entry. Professional licensing requirements, won by private human-service providers, have long been used to shut out competition.

5. *Advocate Vouchers for Low-Income Americans.*

Some critics of privatization maintain that competition—indeed privatization itself—is meaningless in the case of services to the poor, since poor people do not have the financial means to buy adequate private services. Yet vouchers for such services as education, housing, or medical treatment overcome this objection.

6. *Provide Tax Incentives to Encourage Privatization.*

As mentioned earlier, tax incentives concentrate benefits on a small number of people and they act as the nucleus for the growth of privatization coalitions. Incentives are to the emergence of privatization coalitions as new, small spending programs are to public spending coalitions.

Because tax incentives are so essential to a privatization campaign, supporters of the approach should be cautious in their support of tax simplification. Flat tax proposals could be dangerous for privatization because they call for the elimination of deductions as a *quid pro quo* for reduced marginal rates. Privatizers should thus resist the elimination of any deduction that encourages the private provision of a service if otherwise there would be pressure for that service to be provided by government. Eliminating a deduction favoring a private coalition while allowing the equivalent public-spending program to remain on the statute book is like unilateral disarmament in the budget war.

Reversing the Growth of Government

The privatization strategy is essentially incrementalist, or conservative, in nature. While supporters of federal spending rarely neglect an opportunity to create a sweeping new program, spending growth has generally come from a slow but steady stream of small additions to programs. Successful privatization will no doubt occur in the same manner. The privatizer should seek to get his foot in the door with a modest regulatory change, or a small tax break, that will begin the process of coalition building. After that the momentum for growth will develop.

By adopting the privatization strategy, Ronald Reagan could ultimately achieve his objective of achieving a historic reversal in the growth of government, by turning the political flank of the opposition to budget cuts. And by creating permanent coalitions committed to private alternatives to federal spending, the privatization momentum would remain long after President Reagan leaves the White House. But if this administration continues to pursue a head-to-head confrontation with the interest groups, it will be back to business as usual when the Reaganites leave Washington, and Ronald Reagan will merely have won, at best, a short pause in the inexorable expansion of the state.

COMMUNICATIONS

THE NATIONAL BUDGET REFERENDUM: PROCEED WITH CAUTION

Bruce R. Bolnick

In the Fall 1982 issue, Roger A. Arnold and Thomas L. Wyrick[1] (hereafter referred to as A&W) propose a National Annual Budget Referendum (NABR) "to increase the efficiency of the *public sector* at satisfying the desires of the *public*," permitting citizens to "play a more direct role in defining the types of activities that the government is allowed to undertake"(p. 645). A&W propose that a brief report be distributed to the voters each June, explaining the distribution of authorized government expenditures by functional category, together with budget recommendations from Congress and the president for the forthcoming fiscal year. Each voter would then be allotted 100 "expenditure points" to allocate by functional category, or to the category "Refund." Aside from interest obligations on Treasury debt, Trust Funds, or special expenditures authorized by a two-thirds vote of both houses of Congress, the pattern of government expenditures would then be determined by the voters through their allocation of expenditure points.

A&W conclude that the NABR "is a fairly costless move in the direction of direct democracy . . . in an area where reform is most needed"(p. 652). With declining public confidence in government, deteriorating tax compliance, and the frustrating inability of the pub-

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

The author is Assistant Professor of Economics at Northeastern University.

[1] Roger A. Arnold and Thomas L. Wyrick, "Budgetary Referenda: An Efficient Alternative to Representative Democracy," *Cato Journal* 2 (Fall 1982): 637–52. Page numbers shown in parentheses in the text refer to this source.

lic to express its preferences through the ambiguity of the ballot box, it would indeed be highly desirable to find some means to enhance the degree of public control over government budgets. However, the NABR proposal is fraught with serious problems that caution strongly against instituting such a referendum as a binding determinant of budget allocations.

To begin with, the NABR would shift power to the voters only with regard to the broad allocation of the budget by functional category. The nuts and bolts of every tax and expenditure decision would remain beyond their direct control. As a result, the "binding" referendum could easily be circumvented. Broad functional categories are highly susceptible to fiscal gerrymandering; who would doubt, for example, that determined politicians could rationalize food stamps under the allocation for defense if pressed to do so? Also, Trust Funds, which are exempted by A&W from the NABR, could proliferate. Straightjacketing a sandpile can be problematic.

Suppose, though, that the NABR were to succeed in constraining fiscal decisions. The result could be far from wholesome. If final budget allocations were to be fixed by the rigid arithmetic of a referendum, much of the serious and complex work of debate and compromise on the budget nuts and bolts would have to be compressed into the brief period after referendum results were tabulated—hardly a blueprint for good government.

Similarly, administration of authorized programs, and compliance with statutory commitments could be impaired (barring a two-thirds vote of Congress). Granted, programs "ought" to grow or shrink in line with preferences of the electorate. A binding referendum, however, leaves no buffer to control for aberrations in poll results, swings in the public mood, or within-year changes in circumstances. To the voter sitting at home fiddling with the figures, a 1 or 2 percent reallocation of expenditures—or even 5 percent—may be treated very casually. Yet on such budget whimsy could swing tens of billions of dollars of funds. It is easy to imagine a NABR disrupting the administration of veteran's benefits or the judiciary, or alternatively flooding these activities with excess funds. Inconsistencies can be especially acute in the case of entitlement programs involving obligations that fluctuate with the condition of the economy.

Serious problems would arise also in the form of inefficient policy designs provoked when NABR restrictions conflict with political exigencies. For example, farmers' incomes can be supported through surplus purchases, or government-imposed cartel quotas. Unemployed steel workers can be assisted through retraining programs or restrictions on imports. Water quality targets can be met through government programs or through regulations imposed on the private sector. Often the policy options that conserve on budget funds impose the greatest social costs. With a binding NABR the system would lose degrees of freedom in choosing *how* to skin the policy cat, with potentially perverse results.

Strategic voting problems would introduce another arbitrary element.[2] Suppose that one individual's true preferences were those shown in column A of Table 1. If this person expects the actual vote will be something like that in column B, he may very well choose to vote as in column C, despite the fact that he would be horrified to find column C actually adopted.

Now consider the community of five people (I through V) shown in Table 2, with person I being the one whose strategic voting decision has just been examined. The NABR would result in the budget allocation shown in the "Average" column even though *every* individual prefers less to be spent on natural resources and environment. In the extreme, strategic voting could result in no funds at all for defense if everyone believed everyone else would support it. Though contrived, these cases serve to demonstrate the dubious normative significance of the budget referendum.[3]

TABLE 1

AN EXAMPLE OF STRATEGIC VOTING

Program	Expenditure Allocation (in points)		
	A	B	C
Defense	30	35	0
Natural Resources and Environment	10	5	100
Education	15	15	0
Health	15	15	0
Income Security	30	30	0
Refund	0	0	0
Total Allocation	100	100	100

NOTE: A = true preferences of voter in question; B = voter's expectation of referendum result; and C = voter's strategic allocation of expenditures.

---

[2] See Dennis C. Mueller, *Public Choice* (Cambridge: Cambridge University Press, 1979), pp. 56–58, 198–9 regarding strategic misrepresentation of preferences.

[3] A&W discuss the strategic voting problems which can occur in a three-way, winner-take-all election, with the result that such elections "may provide false or misleading signals to policymakers"(p. 641). However, they do not recognize that the same problems will distort the results of the NABR. Indeed, under their proposal policymakers would not just be "misled," but would be bound to the results of the referendum.

TABLE 2

EXPENDITURE ALLOCATIONS FOR A FIVE-VOTER COMMUNITY

Program	Voter I	Voter II	Voter III	Voter IV	Voter V	Average
Defense	0	20	30	50	25	25
Natural Resources and Environment	100	20	5	0	5	26
Education	0	20	20	15	10	13
Health	0	20	20	15	10	13
Income Security	0	20	25	20	10	15
Refund	0	0	0	0	40	8
Total Allocation	100	100	100	100	100	100

Another consideration is that with each vote being "worth" $4000 or more in terms of control over government expenditure, there is considerable scope for abuse of a mail-in referendum. An active market would arise in which those who stand to benefit from government programs endeavor to accumulate ballots in exchange for cash payments.

On a more technical level, the NABR as proposed entails a hidden bias toward reducing the size of government—a result which may be desired, but hardly in accord with the premise of direct democracy. Specifically, by designating expenditure points for "Refund," voters can opt to reduce government expenditures. However there is no symmetrical device for a voter to opt for an increase in the overall scale of government activity. When votes are averaged, there will *always* be a decline in total government spending as long as there is *anyone* who wants a Refund. In Table 2, four of five people prefer no reduction in total expenditures, yet voter V manages to mandate an 8 percent reduction in spending. If these preference patterns persisted, this result would recur year after year. After 10 years total spending would have declined by 56 percent, against the will of 80 percent of the electorate.

Even the seemingly unobjectionable stipulation that "all eligible voters" be allotted 100 expenditure points turns out to cause problems. With the distribution of tax obligations being very skewed in comparison to the distribution of voting rights, a direct budget referendum in which each voter counts equally could result in a major reshuffling of power between blocs of voters. If everyone were to vote, the bottom quintile of families would command 20 percent of the budget allocations, while paying less than 1 percent of federal taxes. The NABR could actually aggravate the sense of alienation on the part of those who bear the greatest tax burden. If so, the NABR is no palliative for A&W's concern that Americans are "losing control of their lives and property"(p. 638). For just such reasons, our bastion of democracy has never favored a simple democratic decision rule for multilateral agencies such as the International Monetary Fund or the World Bank.

On the other hand, if the exercise becomes too complex the NABR could end up being an instrument of elitist, not populist control. Adopting the broadest classification scheme used by the Office of Management and Budget, voters would be confronted with 16 broad functional categories of government activity. If, for the sake of greater precision, expenditures were disaggregated one step to the level of major program categories, fully 70 options would vie for each voter's attention.[4] Faced with a list of bewildering variety, or with a limited number of extremely broad options, many eligible voters might shun the whole process. After all, each individual's preferences would have a weight of roughly one one-hundred-millionth.[5] And many of those who do vote might do so carelessly for the same reasons.

Moreover, given that the list of categories cannot be elaborated too finely, some of the sorest spots in the present budget process would remain unremedied by the NABR. A major source of dissatisfaction involves program-level allocations rather than broad functional allocations of budget dollars. For example, people outraged at government spending on abortions may not wish for less spending on health care services. The more aggregated the referendum format, the more limited the benefits of reform.

Thus, although the NABR proposal points in an attractive direction, the architecture of reform needs extensive rethinking. Fundamentally, the budget determination system must be far more flexible. It might appear that increased flexibility conflicts with the objective of the reform, which is to shift power away from the representatives. What is most important, though, is to strengthen, from a very low base, the effective degree of public control over the scope and structure of government activities—not to give the voters total control. In a world of complex and interactive issues and interests, there are benefits to having representatives who specialize in policy analysis, deliberation, compromise and debate of the sort that occurs within Congress. Hence a more limited referendum might be most desirable.

---

[4] See *The United States Budget in Brief: Fiscal Year 1984* (Washington, D.C.: Executive Office of the President, 1983).

[5] A&W (on p. 648) argue that if 100 million citizens participate in the budget referendum on over $400 billion of expenditures, each citizen is directing the allocation of $4000 worth of government activity. Hence, they say, the incentive to vote and to do so intelligently is strong. Their arithmetic is right, but not necessarily their conclusion. Suppose I am one of those 100 million voters and that without my vote exactly 25 percent of government spending would go to defense. If I were to vote for no allocation at all to defense, its share would drop to 24.99999997 percent. Some voters may diligently allocate their $4000, but I may be unmoved at the option of influencing directly the tenth decimal place of the budget allocation.

For example, an abstract of the Current Services Budget for the forthcoming fiscal year could be distributed with IRS tax forms each January, accompanied by supporting information and guidance. The referendum could then entail having each voter specify a *reallocation of dollar amounts among categories*, constrained by the consistency condition that the changes add to zero. Voters could be given the discretion of defining changes by aggregates or by program, and permissible votes could cover the gamut from raising spending and taxes, to cutting outlays and the deficit, or simply reallocating funds among spending categories.

For a direct democratic budget vote, the referendum would have to be tabulated with each ballot weighted equally. On the other hand, it would be logical to use weights according to tax payments,[6] if the desire is to return to citizens a greater measure of control over the use made by government of their property. With an advisory, nonbinding referendum, *both* sets of tabulations could be processed. Though budget *control* would remain in the hands of representatives, information from the referendum would greatly enhance public influence over the budget. The key here is the word "influence." The "will of the people," clearly expressed, would provide a strong lever for swinging votes. This sort of an indicative referendum would require no institutional or constitutional change, and, being inherently flexible in its link to the budget, would avoid all of the substantive problems associated with A&W's NABR. Then, as experience with the referendum accumulates, the quality of the results could be monitored to evaluate whether and how the link to the budget could be strengthened.

In summary, it would be highly desirable to find an effective means to increase the degree of public influence over the scope and structure of government activities. But the NABR mechanism proposed by A&W is not the right solution. A more flexible instrument is required, generating regular and systematic information on citizens' budget preferences far superior to any now available. Such information could only enhance the public's influence on congressional action. Indeed, experimentation need not begin at the federal level. Innovative state or local governments can lead the way toward budget reform.

---

[6] In this case the referendum returns could be tied to tax returns. As a by-product, tax compliance could improve if citizens saw a direct link between filing taxes and voting in the budget referendum. Though I am not aware of any test of this hypothesis, it would be remarkable if tax compliance were not affected by frustration and anger over how one's tax dollars are spent. The feasibility of the type of referendum instrument suggested here has been tested and reported in Robert P. Strauss and G. David Hughes, "A New Approach to the Demand for Public Goods," *Journal of Public Economics* 6 (1976): 191–204.

CONTINUING THE DEBATE ON NABR: REPLY TO BOLNICK

Roger A. Arnold and Thomas L. Wyrick

As the owners of government, it is entirely appropriate for voters to participate in the public-choice process unless they choose not to. Our nation's heritage and the history of Western civilization suggest that the idea of *representative* democracy is based more on practical considerations than a philosophical rejection of *direct* democracy. In other places and times (Swiss cantons, New England town meetings) direct methods of public policy making have been utilized with considerable success. Today in the United States, though, most people are too busy and lack the detailed information necessary to run the federal government on a daily basis. From this perspective (which we accept), if some means could be found that avoids or greatly diminishes the practical problems associated with direct democracy, then a move in that direction would be desirable.

In our original article on budget referenda,[1] we proposed a means to achieve greater voter participation in the democratic process: the National Annual Budget Referendum (NABR). Under our proposal voters would allocate funds among broad categories of federal expenditure. We provided a number of reasons why the NABR proposal would solve many of the problems that now exist in the public-choice process.

Professor Bolnick is critical of our NABR model for a number of reasons,[2] though he provides no empirical basis for his criticism. In

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

Roger A. Arnold is Associate Professor of Economics at the University of Nevada-Las Vegas and Thomas L. Wyrick is Associate Professor of Economics at Southwest Missouri State University. The authors wish to thank the Southwest Missouri State University Foundation for funding the NABR survey.

[1] Roger A. Arnold and Thomas L. Wyrick, "Budgetary Referenda: An Efficient Alternative to Representative Democracy," *Cato Journal* 2 (Fall 1982): 637–52.

[2] Bruce R. Bolnick, "The National Budget Referendum: Proceed with Caution," *Cato Journal* 5 (Spring/Summer 1985): 337–43.

reply to Bolnick, we first address his major points of disagreement with our analysis and then offer some empirical evidence in support of our NABR proposal. It seems clear to us that Bolnick's problems with our proposal stem not so much with disagreements of a fundamental nature, but with a basic misunderstanding of the elements of our NABR model.

Bolnick raises the following seven points in his criticism of the NABR proposal.

1. Bolnick worries that the broad functional categories of spending associated with NABR voting could be circumvented. "[W]ho would doubt," he asks, ". . . that determined politicians could rationalize food stamps under the allocation for defense if pressed to do so?" First, if food stamps were presented to voters as a defense activity, then our original proposal to separate government spending into distinct *functional* categories has been breached. Second, and more important, while we agree that some politicians and bureaucrats might try their best to place food stamps in the defense category (to stay with the example), we doubt that they would be successful. This is because those who are pro-defense would likely work hard to prevent the placing of an item under the heading of defense that might anger the expenditure-allocating public.

2. Bolnick claims, "If final budget allocations were to be fixed by the rigid arithmetic of a referendum, much of the serious and complex work of debate and compromise on the budget . . . would have to be compressed into the brief period after referendum results were tabulated—hardly a blueprint for good government." But Bolnick overlooks the fact that under NABR, balloting is the penultimate step in the process, and follows (rather than precedes) debate and compromise. Within our proposal the Congress would (as today) decide which activities have a higher priority than others, and would provide that information to citizens prior to the balloting date. Additionally, elected representatives would indicate to citizens the relative shares of total spending they (that is, the representatives) feel should be devoted to each area of activity. Under these arrangements, we would expect that elected officials would spend more time thinking through and debating the budget under NABR than they do now. Under NABR, unpopular spending programs—if left in the budget—would be more likely to impact other programs (those in the same functional category) than is the case today. For this reason, we believe there would be more thought and discussion relating to the formation of a budget that the public would prefer to "purchase." We hold this to be a move in the direction of better government.

3. Bolnick feels that if NABR closes one door to officials who want to grant political favors, they will find another way of accomplishing the same thing. For instance, he states: "Unemployed steel workers can be assisted through retraining programs or restrictions on imports. Water quality targets can be met through government programs or through regulations imposed on the private sector."

We completely agree. But this is an issue we addressed in our article. We can do no better than refer Bolnick (and others) to the discussion on pages 650-1 for our response.

4. Bolnick seems troubled with the idea that under NABR persons paying little in taxes would have the right to allocate the same dollar amount of government spending as those persons who carry a heavier tax burden. He proposes, instead, a weighted voting scheme. According to him, "... it would be logical to use weights according to tax payments ..." We are less certain of the logic of weighted voting than Bolnick. This country has a long history of one man–one vote; said differently, all citizens have equal ownership shares in their government. Providing for unequal voting rights in NABR (or any other voting plan) would make it unconstitutional—and therefore irrelevant.

5. Besides the weighted voting scheme, Bolnick proposes that a non-binding referendum take the place of a binding NABR. Then, "Though budget *control* would remain in the hands of representatives, information from the referendum would greatly enhance public influence over the budget."

Public opinion polls already provide the non-binding referendums that Bolnick describes. For instance, polls taken over the years show that the American public consistently and overwhelmingly favors a balanced budget. If this has had any influence on public policy (as contrasted with political rhetoric), it is not apparent to us. Bolnick appears to believe that current problems are informational and nothing more. Our belief is that certain difficulties in the public choice process—special interest lobbying, logrolling, and so on—have helped create the problems that now exist, and we think NABR would help remedy some of them.

6. Bolnick criticizes NABR on the grounds that it "entails a hidden bias toward reducing the size of government." By "designating expenditure points for 'Refund,' voters can opt to reduce government expenditures. However there is no symmetrical device for a voter to opt for an increase in the overall scale of government activity." While the latter statement accurately describes the NABR ballot, we do not agree with the conclusion that bias is involved. The total amount of government spending (under our proposal) remains in the hands of

Congress and the president, while NABR voting establishes only the allocation of expenditures. Experience suggests that elected officials do not have a "hidden bias" toward smaller government—so neither does a proposal that leaves control of the total budget in their hands.

The decision to reserve a place on the NABR ballot for a "Refund" vote without a symmetric "Addition" category is justified from another perspective as well. Each option implies an assignment of property rights, and a vote to add to the size of government would give citizens a right that they do not now possess. As owners of government, citizens have the right to limit its size if that is what they desire. But citizens do not have the right to control dollars that are owned privately, and that is what they would have if they were allowed to add on to the government at will. NABR is a plan to give citizens control over the government; it was not designed to give them control over the private economy.

7. Bolnick's most serious charge, and the one to which he devotes the most space, is that NABR would be subject to strategic voting. The scenario he describes is a bleak one, where individuals vote in ways that conflict with their actual preferences and end up with a budget allocation that they would prefer not to be stuck with. The points are illustrated in his Tables 1 and 2.

We are troubled by this part of his discussion, and believe it to be in error. First, there seems to be some confusion about the meaning of strategic voting. When voter I (Table 1) prefers that 30 percent of the total budget be spent on defense and expects that everyone else participating in the referendum will direct 35 percent into that category, he casts zero expenditure points for defense. Bolnick calls this the "voter's strategic allocation of expenditures"—presumably because zero points differs from the "true preference" of 30 points.

But a vote of 0 percent by voter I is not a strategic vote, given his expectation of the way others will vote. It is easy to demonstrate this by example. Suppose that a person's situation is such that he has a preference for one automobile, and he is thinking about buying it. Before he acts, however, a notice comes in the mail informing him that he has just won first prize in a raffle: a new car exactly like the one he thought about buying. Now, if we ask how many cars the individual will buy this year he will say zero. But he is not behaving strategically. He is incorporating into his decision the fact that his automobile needs have already been satisfied. A strategic decision would be one in which a person fails to buy the car he wants in hopes that someone else will notice his lack of transportation and react by buying him the car he wanted in the first place.

Bolnick's voter is in a position comparable to the person who won a car in the raffle. Other voters will (he thinks) provide him with as much defense as he has use for—so why buy more? Bolnick's error is in assuming that a person should behave the same no matter what he expects others to do. This was a point we made in our article by saying (page 649) that "each person's vote is at least partly dependent on the quantities of government goods chosen by others . . . ." Our solution was to provide voters with as much information possible about how others are likely to vote by sampling the public prior to each NABR ballot, and including information on the ballot about presidential and congressional recommendations for allocating funds. In this way each person will know approximately how much of each type of activity is likely to be undertaken without his vote, and he will have a better idea of his own marginal evaluations when it is time to cast a ballot.[3]

So if strategic behavior is the appropriate term to describe what Bolnick has discussed, then the point was contained in our original article rather than discovered anew. Our understanding of the subject is contrary to that presented by Bolnick, however, and rather than say NABR is plagued with strategic voting problems we would say that citizens require a small amount of information before we ask them to vote. Although this is not the place to develop the idea, we have a strong feeling that elected representatives need the same type of prior information, too, before they can be expected to vote reasonably. If this is so, then the criticism has more to do with collective decision making than with the NABR proposal.

Finally, Bolnick offers no empirical refutation of our model. We had hoped that our initial paper would encourage others to do empirical research to see if the NABR proposal was operational. We now have conducted a small-scale NABR referendum among voters in Springfield, Missouri. While the experiment was too limited to draw far-reaching conclusions, our results (in Table 1) are far different from those described by Bolnick. Voters did not elect to allocate zero amounts to the functional categories, either as individuals or as a group. Moreover, there seems to be considerable satisfaction with the current budget distribution; the spending changes implied by

---

[3] A fundamental principal of economics is that a person's marginal utility from consuming a good or service—hence, his marginal evaluation of it—depends upon the quantity available for consumption. It follows that as other voters provide units of (say) national defense, then the amount which one would be willing to purchase for himself will, *ceteris paribus*, decline. This phenomenon is not associated with a particular decision-making mechanism—it is present under prevailing political institutions as well as NABR.

TABLE 1

NABR SURVEY RESULTS[a]

Government Activity	NABR Vote[b]
National Defense	28.8
International Affairs	1.5
Science, Space, Technology	1.9
Energy	1.1
Nat. Resources/Environment	2.0
Agriculture	1.9
Commerce & Housing Credit	0.4
Transportation	2.8
Community & Regional Development	1.0
Educ., Training, Employment & Social Services	4.1
Health	11.5
Income Security	34.0
Veterans Benefits	3.2
Administration of Justice	1.0
General Government	0.7
General Purpose Fiscal Assistance	0.8
Refund to Taxpayers	3.5
TOTAL[c]	100.2

[a] Survey taken in Springfield, Missouri, October/November 1983.

[b] Shares (in percent) of the total budget that those surveyed would prefer to have going into each spending category.

[c] Total does not equal 100 percent because of rounding.

our survey are only marginally different from actual policy. We encourage others to undertake similar surveys and to report their findings. (We would be happy to provide background material for those who request it.)

In conclusion, we continue to believe that representative democracy can be improved upon. Initially, we held that a national annual budget referendum process, philosophically speaking, was the way to go. Today, we believe that there are so few technical problems with the budget referendum process that it stands, more than ever, as an excellent alternative to policy making as we have come to know it. We urge those persons who seek better and more responsive government to seriously research its merits.

BOOK REVIEWS

Political Economy and Freedom
G. Warren Nutter
Indianapolis: Liberty Press, 1983, 314 pp.

This book is a collection of papers by the late G. Warren Nutter, who died in 1979. The essays, many of them previously unpublished, were written mainly during the period from the mid-fifties to the mid-seventies. They have been well chosen to reflect the broad spectrum of Nutter's interests and his contributions to the field of political economy. Most of Nutter's career was spent at the University of Virginia, where he was chairman of the economics department for most of the 1960s. His four-year stint as Assistant Secretary of Defense for National Security Affairs from 1969 to 1973 provided him an opportunity to observe and participate in the "political economy" of U.S. policymaking in the critically important areas of defense and foreign affairs.

The 33 essays are grouped in three parts, each one focusing on a major area of Nutter's interest. The first group—15 papers—reflects Nutter's abiding conviction that human progress and personal liberty are best served by an economic system that maximizes the use of free competitive markets, endorses private ownership of the means of production, and limits the role of government. He also thought that such economic arrangements promote political freedom. As he points out in the essay "I Choose Capitalism," it comes as no surprise that "without exception Communist societies have been monistic and authoritarian, while capitalist societies are typically, though not always, pluralistic and liberal." Several papers are concerned with the relative extent of competition and monopoly in the American economy. These reflect one of Nutter's major research contributions—the finding that the degree of concentration in U.S. industry had not risen since the turn of the century. Other papers are related to his work on the measurement of trends in the size of the government sector and in explaining the reasons for its persistent expansion. Finally, in the essay "Markets Without Property: A Grand Illusion," Nutter stresses the inescapable link between the efficiency (pricing) behavior of a business firm and the nature of the ownership rights it

---

Cato Journal, Vol. 5, No. 1 (Spring/Summer 1985). Copyright © Cato Institute. All rights reserved.

holds, arguing (with Ludwig von Mises) that private ownership provides powerful incentives for economizing on scarce resources. Nutter would not be surprised to learn that would-be reformers in Hungary and Yugoslavia now are grappling with this fundamental issue—how to structure "socialist" property rights so as to induce efficient behavior from business firms. Certainly he would argue that the notion that genuine competitive markets can be created in Communist countries without dissolution of their authoritarian political order is a "grand illusion."

In Part Two we have seven essays dealing with the centrally planned economies in Communist countries. Another of Nutter's major research contributions was his pioneering study of the growth of industrial production in the Soviet Union. One of the essays summarizes the main findings of this study, which, although controversial when first published, have stood the test of time. Thus, Nutter expected Soviet industrial growth to slow secularly, as it has, to rates of under 3 percent annually by Western measures in the early 1980s. Nutter also stressed the peculiar structural pattern of Soviet industrial growth, a pattern "that would never be duplicated in a market economy. Sovereign consumers would not choose the paths of growth chosen by Soviet leaders." Nutter surely would agree that the persistent adherence to such structurally unbalanced growth has contributed importantly to the present malaise in the Soviet economy. Along with the foreign policy ambitions of the Soviet leadership, it also helps to explain why the ordinary Soviet citizen, despite notable gains, is still pretty much "today's forgotten man," just as he was in the 1960s, when Nutter so characterized his lot. Several papers touch on the dilemma of economic reform in the Soviet Union. In words that still ring true in the 1980s, Nutter wrote "Soviet leaders of today face the dilemma of the czars. Like the czars, they know the economy is sick, but also like the czars, they fear that curing the patient will kill the doctor."

Part Three brings together 11 essays, most of which deal with various aspects of U.S. foreign policy (as do some of those in Parts One and Two), particularly its economic aspects. Many of the issues and arguments raised in these papers written in the 1960s and early 1970s are timely today, when the United States must fashion a strategy to deal with difficult global issues made even more complex by growing ferment in the Third World, by allies in Europe and Japan made restive over increased deployment of nuclear arms and frictions over trade, and by a Soviet Union stronger than ever militarily but with a troubled economy and an energetic new leader perhaps proffering another era of detente. While strongly advocating controls on strategic trade with the Soviet Bloc, Nutter speaks of the folly of economic warfare in general and also of the illusion that the United States can "buy" loyalty and friendship through economic and military aid. With respect to superpower relations, he argues against the notion that U.S. policy can bring about internal liberalization in the Soviet Union. Rather, "our policy should be guided by the goal of preserving our own free system. This is a time for hard bargaining, for insistence on a viable quid pro quo in the case of every action we might take to ease Soviet internal problems. . . ." And more generally, "How does a country of peaceful intent make its way through the maze of world politics? . . . It makes sure . . . that it has sufficient strength and resolution. . . . It seeks . . . the path of negotiation as far as other nations will go along. And it prepares itself, finally, to deter and if necessary, to repulse hostilities. To neglect any one of these elements of foreign policy is to invite disaster."

Gertrude E. Schroeder
University of Virginia

Power and Privilege: Labor Unions in America
Morgan O. Reynolds
New York: Universe Books and the Manhattan Institute
for Policy Research, 1984, 309 pp.

The purpose of this book is "to explain and interpret the behavior of unions and unionists—what they do, why they do it, and what effects their actions have . . . in light of contemporary economic analysis." Other labor economists have, of course, attempted this. But in my view none has succeeded as well as Professor Reynolds in cutting through the veil of rhetoric, emotion, and polemics that so often shrouds the activities and effects of unions in contemporary discussion. This is not an "anti-union" or "anti-worker" book: It simply seeks to improve our understanding of what unions do, making use of modern economics, especially developments in public choice and the economics of regulation.

Chapter 1 surveys "the enduring controversy over labor unions," paying particular attention to the role of ideas in the evolution of public policy toward unions. According to Reynolds, the government-sanctioned powers of unions exercised through work disruption, threats of disruption, and other kinds of pressure ultimately rest on some acceptance or tolerance by the general public. This is where intellectuals fit in: They have supplied arguments to make the coercive actions of unions appear reasonable. In particular, the intellectual argument for unions rests on two propositions: (1) individual employees are helpless in dealing with employers, and (2) unions are essential to give employees the ability to deal on an equal basis with an employer. After showing that neither of these views bears close scrutiny, Reynolds concludes that they are based more on "garbled sentimentality toward unions" than analysis.

Chapter 2 begins to present some of the much-needed analysis by discussing the economic nature of unionism. A positive view of unions is first presented which makes a very basic distinction. Namely, in a free society people ought to be able to associate freely with one another, to form "combinations of labor," as long as they do not interfere with the equal rights of others. Moreover, people ought to have the right to quit or avoid such associations as well. Thus, there is nothing inherently objectionable about multiperson organizations called unions, and unions can often improve social coordination, at least in theory.

In reality, however, the practice of unionism is anything but conducive to a free society. Unions are essentially labor cartels which seek to "take competition out of wages," to use a classic union slogan. As Arthur J. Goldberg, former general counsel of the AFL-CIO once wrote, "any labor union is a monopoly"; "it eliminates competition between employees for the available jobs." Strikes have long been heralded as the major "weapon" in the union "arsenal," and, as Reynolds points out, the violence surrounding strikes exposes the basic conflict in the labor market, namely, the conflict between organized and unorganized labor, not "labor" and "management." Reynolds views strikes as a breakdown in the rule of law: "Hitting a person over the head with a baseball bat to take $20 from his wallet is a crime on the street, but it is much less likely to be treated as criminal if the person wielding the bat is an organized [that is, unionized] worker in a labor dispute . . ." (p. 50). If the rule of law prevailed, says Reynolds, unions would not be able to coerce anyone. As the author demonstrates throughout the book, the long-run answer to monopoly unionism is to eliminate the special legislation, legal immunities, and other government empowerments and treat unionists like everyone else under tort and contract law.

Chapter 3 surveys a number of myths that have evolved and have helped unions achieve their special governmental empowerments. One myth is that "unions protect workers from employers' superior bargaining power." Reynolds shows that this belief is almost entirely without basis in theory or fact, and is largely a result of union propaganda, distortions emanating from the academic community, and the "rational ignorance" of the public. Another myth holding that "high living standards in the U.S. are due to a strong union movement," is also exposed. Countries with very militant unions, such as England, have experienced wage growth far slower than in the United States, and there is much evidence that by impairing productivity unions actually impoverish a nation. Besides, if unions caused high living standards, one would have to accept the preposterous notion that poverty in say, India, could be eliminated by rapid unionization. The "company is the enemy" myth is also easily disposed of by recognizing that unions do not compete with employers, but rather with those who sell substitutes for their members' services—rival unions, foreign workers, nonunion workers, labor-saving and cost-reducing machinery, and so on. The history of bitter, interjurisdictional disputes among unions is evidence of this, as is their historic opposition to free international trade. Finally, the notion that "unions protect workers against the abuse of managerial authority" is also dismissed as an unsubstantiated myth. Economic common sense would indicate that competitive labor markets are the ultimate protectors of employees, not union bureaucracies. Employers who fail to listen to employees' grievances will have to pay higher wages to attract workers in competitive markets, which ultimately leads to lower profits and possible bankruptcy.

Chapter 4 discusses the latest collectivist rationale for compulsory unionism, the "Harvard–National Bureau of Economic Research" view, as developed by two Harvard economists, Richard B. Freeman and James L. Medoff. The Freeman-Medoff view of unions, in a nutshell, is that unions are a vehicle for collective voice—for providing workers with a means of communicating at the workplace and in the political arena. Consequently, they generally increase productivity, promote economic equality, and are said to be democratic, noncorrupt organizations. This view is directly at odds with the alternative view (held by Reynolds) that unions are labor monopolies. In criticizing this approach to the study of unionism Reynolds correctly points out that the Harvard School simply ignores the past 20 years of research in public finance and public choice. It is no longer "legitimate" (not that it ever was) for economists to claim that a market that does not pass the test of some utopian ideal, for example, Pareto optimality, provides a necessary and sufficient condition for government intervention. Even if markets are not "perfect," government intervention has a long record of making matters worse, not better, but this is ignored by the Harvard School. Freeman and Medoff claim that there are public goods and externality problems in the workplace that require government intervention in favor of unions, but ignore the fact that imposing one exclusive union bargaining agent on all workers will *create* externalities by benefiting some unionists at the expense of other workers. Moreover, their own data show this. For example, they find that unions reduce the dispersion of wages within unionized industries, indicating that less productive workers gain at the expense of more productive workers whose wages are dragged down toward the average "union rate."

Reynolds also finds inconsistencies in the productivity-enhancing views of the Harvard School and, finally, points out that unions are anything but democratic: They are basically autocracies. Less than 10 percent of American unions actually hold elections for their leadership, and in those which do, the candidates are usually chosen by executive boards whose preferences are rarely challenged effectively.

Chapter 5 is a summary of many of the laws and regulations that render unions privileged organizations in American society, and which give them special advantages over other citizens and groups. It is shown how unions have almost always depended on special legislative privileges for, like all other cartels, market forces render them unstable over time. Union reliance on special legislative treatment is demonstrated in a thorough discussion of the Norris-LaGuardia Act, the Wagner Act, and other depression-era legislation. Reynolds abolishes the myth that these laws were depression-era responses to "market failures," as has long been argued by academics, and shows how the laws were passed to create monopoly power in the union representation business. They were not "public interest" laws, but laws that promoted the private self-interests of unionists, politicians, bureaucrats, some academics, and even a few businessmen.

Chapters 6 and 7 survey the economic effects of unionism, while chapter 8 discusses the "new unionism in the public sector." In this chapter we are shown how, as with private sector unions, special legislative treatment has contributed to the power of public employee unions. This special treatment enables their members to earn as much as 40 percent more than equally qualified private sector workers, not counting the superior pensions, fringe benefits, and working conditions that often exist in the public sector.

Chapter 9 sets the record straight on the relation between unionization and inflation. Although unionization may affect *relative* prices, it does not necessarily have anything to do with the cause of a continual, sustained increase in the price *level*. Monetary growth in excess of the growth of real GNP is the cause of inflation. The subtle point about the role of unions in the inflationary process is that they, and their interest-group allies, are strong lobbyists for expansionary governmental policies that ultimately lead to the monetary growth that causes inflation.

Just when you'd thought you'd heard it all, chapter 10 discusses even more "contradictions of unionism." This may sound a bit repetitive to the reader of this review, but since unions, politicians, and academics have spent the past half-century spreading misinformation and propaganda about unions, several other books could easily be written on the subjects covered in chapter 10. Here it is shown that, contrary to popular mythology, unions have long been anti-free enterprise (look at their support of an interventionist "industrial policy," for instance); they are a major source of racial discrimination (W. E. B. DuBois called them "the greatest enemy of the black working man"); corruption is a major feature of unionism (ask anyone in the construction business); and unions promote political corruption and instability.

The final chapter concludes that unionism is "a subset of the man-made restrictions on human action." In short, unionism is an important element of the collectivist movement in America.[1] Reynolds discusses the future of unionism which includes a trend toward what has been called "pushbutton unionism," union reliance on pressure directed at employers and government, rather than employees, to keep members.[2] Finally, the author points to a way out of the problems created by unionism: restoring equality before the law. This is not to suggest "anti-union" or "anti-worker" legislation, but simply to repeal the legislative privileges that have enabled unions to earn monopoly profits by the use of force and threat at the expense of everyone else in society. More specifically:

---

[1] For a more detailed discussion of the role of labor unions in the collectivist political movement in America, see James T. Bennett and Thomas J. DiLorenzo, *Destroying Democracy: Government Funding of Partisan Politics* (Washington, D.C.: Cato Institute, 1985).

[2] James T. Bennett and Thomas J. DiLorenzo, *Labor Unions and the State: The Political Economy of Pushbutton Unionism* (Dallas: The Fisher Institute, 1985).

> [T]he right thing to do is to deregulate. Try freedom for a change. Repeal, abolish, rescind, revoke, and do away with the Railway Labor Act, the Norris-LaGuardia Act, and the National Labor Relations Act. Also abolish the commissions, executive orders, state laws, rulings, administrative orders, and regulations derived from the three major statutes. Restore the rule of law in labor relations by treating unions in a manner consistent with the way everyone else is treated under contract and tort law. Treat workers and worker organizations as responsible adults, not as children who are exempt from the rules of peaceable conduct. [p. 264]

If more people became aware of the truths about unionism that Professor Reynolds has revealed, the prospects for greater freedom would surely improve.

Thomas J. DiLorenzo
George Mason University

The Theory of Public Choice—II
Edited by James M. Buchanan and Robert D. Tollison
Ann Arbor: University of Michigan Press, 1984, 512 pp.

This is the second volume of papers devoted to public choice theory edited by James Buchanan and Robert Tollison. Twelve years elapsed between the 1972 publication of the original *Theory of Public Choice* and the current volume. Comparison of the two volumes (hereafter referred to as PC–I and PC–II) gives some insight into how public choice has developed in the intervening years. This perspective is somewhat blunted by a disproportionate representation of the "Virginia school of political economy," especially in PC–I. Because the Virginia school has been the source of so many contributions to public choice theory, however, the parochial selection process does not seriously detract from the value of the two volumes.

In his introduction to PC–I, Tollison concluded that "[t]he 'theory of public choice' remains on the threshold of development." A theme of his preface to PC–II is that public choice is well past the threshold of acceptance as a legitimate subdiscipline in economics. He notes the growth of the journal *Public Choice* and also that public choice articles are now more widely published in major economics journals than was the case in the 1960s and early 1970s. A comparison of original publication outlets for the papers in PC–I and PC–II supports this observation. In addition, textbooks in public finance routinely include material on public choice. Students are also receiving exposure to the field in some major elementary textbooks.

The first two papers in PC–II, following Tollison's introduction, provide an excellent overview of public choice. James Buchanan's "Politics without Romance . . ." is a concise summary of the origins of public choice, its methodology and perspective on the workings of government. Public choice has profoundly altered the way modern economists approach public policy analysis. The stance of proffering advice to a benevolent despot—of searching for a set of policies which would allow that despot to maximize social welfare—is increasingly abandoned as irrelevant. Buchanan outlines the alternative perspective of public choice—that of viewing the public sector as a set of institutions through which individuals seek to achieve their goals. The tools and methods which economists have traditionally and successfully applied in their analysis of market institutions are extended to the analysis of political institutions. The second paper is Dennis Mueller's excellent survey article originally published in the *Journal of Economic Literature*. Where Buchanan offers a general outline, explaining what public choice is and why it is important, Mueller fills in the details of how public choice scholars go about their work. Both papers should be required reading for anyone beginning a study of the field.

Voting theory has always been an important part of public choice. The neglect of this area in PC–I has been remedied in PC–II by the inclusion of several papers. The most important is Nicolaus Tideman's and Gordon Tullock's paper on "A New and Superior Process for Making Social Choices." Their "demand-revealing process" is a major contribution to literature dealing with the problem of preference revelation in non-market settings. Other useful and provocative papers include Dennis Mueller's "Voting by Veto," a paper by Mueller, Tollison, and Thomas Willett "On Equalizing the Distribution of Political Income," and Peter Bernholz, "A General Social Dilemma: Profitable Exchange and Intransitive Group Preferences."

PC–I was devoted almost exclusively to policy analysis. This was appropriate in that public choice scholars needed first to challenge mainstream economics in its traditional domain and to demonstrate the valuable insight the public choice perspective lends to policy questions. The repeated demonstration of that insight undoubtedly played a major role in taking the romance out of politics, to use Buchanan's phrase. Policy analysis remains important. In the mid-1980s, however, a major new area of inquiry has come into its own. This is the economic theory of constitutions. If policy analysis can be characterized as positive analysis of how individual and group decisions on specific issues are affected by existing political institutions, the theory of constitutions takes a step back and examines the choice of institutions themselves. Building on positive analysis of how institutions affect outcomes, constitutional economics addresses itself to the normative issue of which institutions are best suited to achieving desired results.

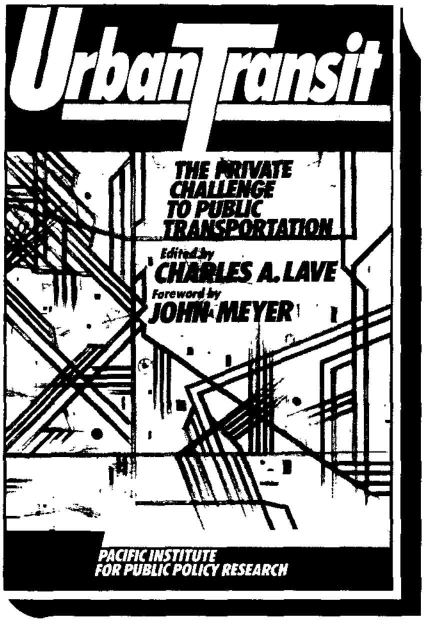
Here, as in so many areas of public choice, much of the intellectual leadership has come from James Buchanan. In addition to the discussion in his introductory paper, Buchanan has authored or coauthored three other papers in PC–II which deal with constitutional economics. The concluding paper in the volume, "Constitutional Restrictions on the Power of Government," is essentially a discussion of how constitutional economics relates to important issues confronting the American polity, in particular, proposals to amend the national and various state constitutions to restrain government's power to tax and spend. Two other papers, coauthored with Geoffrey Brennan, demonstrate in more depth the intellectual logic of constitutional economics. "Towards a Tax Constitution for Leviathan" addresses the issue of appropriate tax norms or constitutional constraints which might be desirable if majority voting fails to restrain the revenue-maximizing productivities of "Leviathan" government. Stated succinctly, the Brennan-Buchanan analysis stands traditional normative tax theory on its head.

The Leviathan model has been controversial, primarily because of the unwillingness of many scholars to accept its empirical relevance. Buchanan and Brennan provide a provocative argument in support of their approach in another paper, "The Normative Purpose of Economic Science." Although this paper appears in a different section of PC–II than "Towards Tax Constitution. . . ," it is strongly recommended that these two papers be read together.

PC–II contains other important contributions to public choice theory which have not been mentioned here due to space constraints. The volume is generally excellent and is a valuable source for anyone interested in the field.

Marilyn R. Flowers
University of Oklahoma

THE AUTHORITATIVE NEW BOOKS


Urban Transit
THE PRIVATE CHALLENGE TO PUBLIC TRANSPORTATION
Edited by CHARLES A. LAVE
Foreword by JOHN MEYER
PACIFIC INSTITUTE FOR PUBLIC POLICY RESEARCH

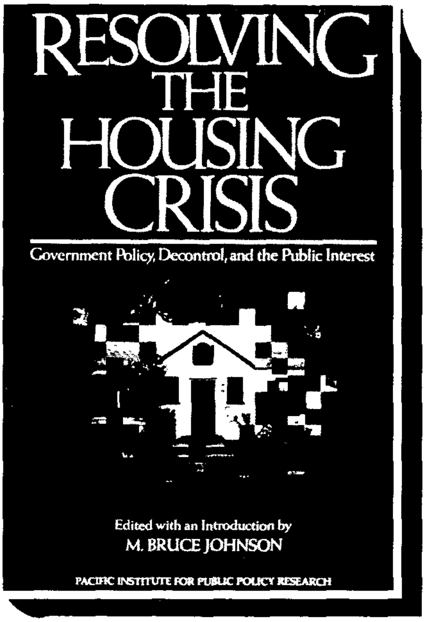
URBAN TRANSIT
The Private Challenge
to Public Transportation
Edited by CHARLES A. LAVE
Foreword by JOHN MEYER

Decades of active government involvement in municipal transportation have left many transit systems on the brink of financial collapse, encouraged inefficient service and have insulated management from the discipline of market forces. Consequently, *Urban Transit* focuses on the economics of transit matters, the possible alternative modes of urban transit, and the economic and political problems of implementing changes. Concluding that private, free, competitive market forms of transit would solve these problems, *Urban Transit* is a sourcebook of workable remedies to effect consequential and necessary changes in this increasingly crisis-prone social dilemma.

"Urban Transit *is extremely interesting and indeed a landmark.*"
–MARTIN WACHS
Professor of Architecture and Urban Planning UCLA

"*This is a timely book of profound importance to anyone interested in the future of urban transportation.*"
–RALPH L. STANLEY
Urban Mass Transportation Administration

ca. 354 Pages • ISBN 0-88410-970-4 Paper $12.95
ISBN 0-88410-969-0 Cloth $30.00


RESOLVING THE HOUSING CRISIS
Government Policy, Decontrol, and the Public Interest
Edited with an Introduction by
M. BRUCE JOHNSON
PACIFIC INSTITUTE FOR PUBLIC POLICY RESEARCH

RESOLVING THE HOUSING CRISIS
Government Policy, Decontrol,
and the Public Interest
Edited with an Introduction
by M. BRUCE JOHNSON

Across the United States, the problems of housing costs and availability have seldom before been so acute: home ownership has become virtually inaccessible to the average person; construction costs are high and rising; and government restrictions increasingly prevent real estate growth. *Resolving the Housing Crisis* is a comprehensive treatment of the increasingly serious problems in housing and construction. This volume analyzes the underlying contributing factors to the crisis, the responses in land use, housing, and construction policies, and the interaction between the crisis and the legal status of property rights. As a result of this assessment, the authors argue for substantial deregulatory housing and land use reforms.

"Resolving the Housing Crisis" *is a first-rate job. . . . One of the most inclusive contributions to the literature I have ever had the pleasure of reading.*"
–GEORGE STERNLIEB
Director, Center for Urban Policy Research
Rutgers University

440 Pages • ISBN 0-88410-386-2 Paper $13.95
ISBN 0-88410-381-1 Cloth $26.00

To order, or for a catalog of publications, please write:
PACIFIC INSTITUTE, 177 Post Street, Dept. C4
San Francisco, CA 94108

PACIFIC INSTITUTE
FOR PUBLIC POLICY RESEARCH

International, authoritative, controversial . . .

JOURNAL OF HEALTH POLITICS, POLICY AND LAW

Founded in 1976 to provide a forum specifically dedicated to communicating information and ideas about health policy, the *Journal of Health Politics, Policy and Law* is now widely acknowledged as the most notable periodical in a field it has helped to create. Its scope is interdisciplinary: contributors in 1984 included lawyers, medical practitioners, government officials, health insurers, and academics from such fields as business administration, urban planning, gerontology, and religion, as well as the standard array of social scientists and policy analysts.

Recent Articles Have Included:

John Holahan
State Rate-Setting and its Effects on the Cost of Nursing-Home Care

Richard B. Saltman
Power and Cost Containment in a Danish Public Hospital

Frank A. Sloan
State Responses to the Medical Malpractice 'Crisis' of the 1970s: An Empirical Assessment

Muriel R. Gillick
Health Promotion, Jogging, and the Pursuit of the Moral Life

Each issue also contains an extensive Book Section and a News and Notes Section providing information across the disciplines on achievements, conferences, grants, new publications, and calls for papers.

Please enter a one-year subscription to the *Journal of Health Politics, Policy and Law* (4 issues) at the rate checked below:

_____ $60/Library or Institution ☐ Payment enclosed
_____ $36/Individual (by personal check only) ☐ Please bill me
_____ $16/Student

NAME: ____________________

ADDRESS: ____________________

____

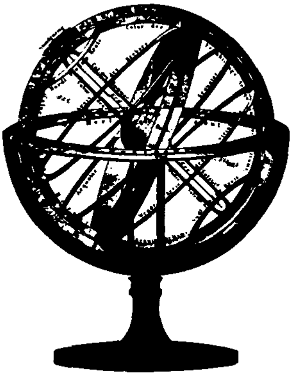
MAIL TO:
Box 3018, Duke University Medical Center, Durham, NC 27710

INTERNATIONAL ATLANTIC ECONOMIC CONFERENCE

INVITATION FOR PAPERS

APRIL 15 - 20, 1986

St. Thomas
Virgin Islands



"CHARTING NEW FRONTIERS"

SUBMISSION DEADLINE: October 15, 1985

Author Submission Fee

All authors wishing to present a paper should include a submission fee of $35.00 (U.S.) for each paper. To avoid a bank collection charge, all foreign and Canadian checks should be drawn on a U.S. bank. Submission fees are not returnable. If there is more than one author per paper, only one submission fee is required.

Authors

Those wishing to have a paper considered for acceptance must submit two copies of a 500 word summary, a submission fee, and separate cover sheet listing: (1) Location of conference to be considered for; (2) Name of author(s); (3) Institution of affiliation; (4) Mailing address of author(s); (5) Phone number of author(s); (6) Number and name of the category under which the article primarily belongs.

Discussants and Chairmen

Those preferring to be a discussant or chair a session, please submit the following information: (1) Location of conference to be considered for; (2) Name; (3) Institution of affiliation; (4) Mailing address; (5) Phone number; (6) Preference for discussant, chairman, or both; (7) Number and name of your primary and secondary interests. Chairmen and discussants are not required to pay a submission fee.

All participants accepted on the program will be responsible for their own expenses and payment of the conference registration fee.

Submit to: John M. Virgo, Atlantic Economic Conference, Southern Illinois University
at Edwardsville, Box 101, Edwardsville, IL 62026-1001, U.S.A.

> “It’s the very best journal of its kind. It is substantial, relevant, and a must for policy analysts, practitioners, and academicians.”
>
> —Amitai Etzioni, The George Washington University

Journal of Policy Analysis and Management

Editor: Raymond Vernon, Harvard University

The JOURNAL OF POLICY ANALYSIS AND MANAGEMENT *is the official journal of the Association for* Public Policy Analysis and Management.

Published quarterly, the **JOURNAL OF POLICY ANALYSIS AND MANAGEMENT** represents a vital international forum for the publication of outstanding contemporary research and analysis in the field. As a subscriber, you’ll find incisive, original papers on every major area of domestic and international public policy, as well as articles on how technology and important analytical tools can further sharpen your policy-making skills. Reading JPAM, you’ll engage in the mutual exchange of ideas with policy analysts, academicians, government managers, and corporate officials from all over the world. And you’ll better define your own professional identity and role with the help of JPAM’s cogent articles on ethics, tactics, and the inevitable dilemmas that confront policy-makers.

Recent issues have included such in-depth research as:

- The Congressional Veto: Shifting the Balance of Administrative Control—*Robert S. Gilmour*
- Professionals at Bay: Managing Boston’s Public Schools—*Robert Wood*
- Is the Consumer Price Index a Fair Measure of Inflation?—*Richard W. Wahl*
- International Agencies and the Art of the Possible: The Case of the IEA—*Robert O. Keohane*
- Government Support of Technical Progress: Lessons from History—*Richard R. Nelson*
- Restraining Auto Imports: Does Anyone Win? —*Jose A. Gomez-Ibanez, Robert A. Leone and Stephen A. O’Connell*
- The Bird in the Hand: A Feasible Strategy for Gun Control—*Mark H. Moore*

Supported by such regular features as topical book reviews, listings of working papers on ongoing research, and case notes for public policy instructors, the **JOURNAL OF POLICY ANALYSIS AND MANAGEMENT** is your source for the most complete perspectives on advances in this rapidly moving field.

JOHN WILEY & SONS
605 Third Avenue
New York, NY 10158

Send to:
JOHN WILEY & SONS, 605 Third Avenue, New York, NY 10158

Please enter a one-year subscription to the JOURNAL OF POLICY ANALYSIS AND MANAGEMENT (4 issues).

☐ Individual $36 ☐ Institutional $44
☐ Special Discount for individual ASPA members $30
☐ Bill me. ☐ Payment enclosed.
Charge to: ☐ MasterCard ☐ VISA ☐ American Express

Card No.____________ Expiration Date______

Signature ____________

For subscriptions outside U.S. add $19 for surface postage, or $84 for airmail delivery. Payment for international orders must be in U.S. currency. Service begins when payment is received.

Name____________
Address____________
____
City/State/Zip____________

TO ORDER BY PHONE
CALL TOLL FREE 1-800 526-5368.
IN NEW JERSEY CALL COLLECT 201 342-6707.

5-1841

Reason Papers

A Journal of Interdisciplinary Normative Studies

Articles

Is Economics Independent of Ethics? . . . . *Jack High*

How the Jacksonians
Favored Industrialization . . . . *Paul McGouldrick*

Economic Warfare
in Defense of Liberty . . . . *Juliana Geran Pilon*

Choice and Rationality . . . . *Antony Flew*

On Improving Mankind
by Political Means . . . . *Lester H. Hunt*

Constraining the Choice Set: Lessons
from the Software Revolution . . . . *David Levy*

Discussion Notes

The Randian Argument Reconsidered:
A Reply to Charles King . . . . *Paul St. F. Blair*

Is "Flourishing"
a True Alternative Ethics? . . . . *David L. Norton*

Review Essays & Reviews

John Gray's *Mill on Liberty: A Defence*
and *Hayek on Liberty* . . . . *Nicholas Capaldi*

Tom Regan's *The Case for
Animal Rights* . . . . *John Hospers*

REASON PAPERS is published by the Reason Foundation, 1018 Garden Street, Santa Barbara, CA 93101. Send orders ($5.00 per copy in US and Canada, $6.00 elsewhere) to Reason Papers, Box 40105, Santa Barbara, CA 93140. Manuscripts should be accompanied by return postage and envelope.

Copyright © 1985 by the Reason Foundation. All rights reserved.

NO. 10 **SPRING 1985**

Zeitschrift für die gesamte Staatswissenschaft

Journal of Institutional and Theoretical Economics (JITE)

Symposium Issue

Vol. 141 March 1985 No. 1

Barzel, Yoram
Transaction Costs:
Are They Just Costs?
Comment by Bruno Frey

Neumann, Manfred
Long Swing of Economic Development, Social Time Preference and Institutional Change
Comment by Eirik G. Furubotn

Summers, Robert S.
Some Considerations Which May Lead Lawmakers to Modify a Policy When Adopting it as Law
Comment by Warren J. Samuels

Behrens, Peter
The Firm as a Compley Institution
Comment by Kenneth E. Scott

Hax, Herbert
Economic Aspects of Bankruptcy Law
Comment by Victor P. Goldberg

Crew, Michael and Kleindorfer, Paul
Governance Costs of Rate of Return Regulation
Comment by Erich Kaufer

Schmid, Allan A.
Intellectual Property Rights in Bio-Technology and Computer Technology
Comment by Arnold Picot

Frech III, Harry E.
The Property Rights Theory of the Firm:
Some Empirical Evidence from the U.S. Nursing Home Industry
Comment by Gérard Gäfgen

Seyfert, Wolfgang
A Weberian Analysis of Economic Progress:
The Case of Resource Exporting LDCs
Comment by Armen A. Alchian

Williamson, Oliver E.
Reflections on the New Institutional Economics

Editor: Rudolf Richter
Editorial Board: Armen A. Alchian, Eirik G. Furubotn, Terence W. Hutchison, Heinz König, Ernst-Joachim Mestmäcker, Wolfgang F. Stolper.

Papers for Publication in English should be sent in triplicate to:
Professor Dr. Rudolf Richter, D-6600 Saarbrücken 11, Universitätsgebäude Nr. 31, W. Germany.

J.C.B. Mohr (Paul Siebeck) Tübingen

Editors:

Mogens Pedersen, *University of Odense, Denmark*
Derek Urwin, *University of Warwick, England.*

Editorial Board

Dag Anckar, *Abo, Finland*
Francis G. Castles, *Milton Keynes, England*
Hans Daudt, *Amsterdam, The Netherlands*
Gunnel Gustafsson, *Umeå, Sweden*
Hans Dieter Klingemann, *Berlin, FRG*
Gianfranco Pasquino, *Bologna, Italy*
Annick Percheron, *Paris, France*
Geoffrey K. Roberts, *Manchester, England*
Julian Santamaria, *Madrid, Spain*
Manfred G. Schmidt, *Konstanz, F.R.G.*
Marie-France Toinet, *Paris, France*

1985: Vol. 13 (4 issues)
For libraries and institutes
US $ 73.00/Dfl. 197.00, including postage
For individuals
US $ 33.25/Dfl. 90.00, including postage
Back volumes are available on request

Ask for a FREE SAMPLE COPY

EUROPEAN JOURNAL OF POLITICAL RESEARCH

Official Journal of the European Consortium for Political Research

Unlike other journals the *European Journal of Political Research* does not exist merely as another scholarly publication.
Its principal objective is still to foster communication and collaboration about ongoing research across national and linguistic boundaries within Europe. Its programme for the 1980s now includes even more stimulating articles, more book reviews, and more news on current political research - essential reading for all those interested in political science.

Some recent articles:

On the Futility of the 'Demand Approach' to Public Sector Growth, **Ole P. Kristensen**
Political - Economic Interdependence and the Management of Economic Crisis, **Franz Lehner and Hans Keman**
The Intergenerational Transfer of Political Ideologies in Eight Western Nations, **M. Kent Jennings**
General Elections in Western Nations During 1983, **Thomas T. Mackie and Richard Rose**
Normal Vote Analysis: A Methodological Note, **Hans Rattinger and Jürgen W. Falter**

ELSEVIER

NON SOLUS

P.O. Box 330,
1000 AH Amsterdam,
The Netherlands.

52 Vanderbilt Ave.,
New York, N.Y. 10017

The Dutch guilder price is definitive. US $ prices are subject to exchange rate fluctuations.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

☐ DIALOG/Dialorder
☐ OnTyme
☐ ITT Dialcom
☐ OCLC ILL Subsystem
☐ Other (please specify)__________
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name__________
Title__________
Institution/Company__________
Department__________
Address__________
City__________State_____Zip__________
Phone (_____)__________

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

Southern
Economic
Journal

Joint Publication of the
Southern Economic Association and the
University of North Carolina

The *Southern Economic Journal* is published four times a year, in January, April, July, and October, at Chapel Hill, N.C. 27514. The subscription price is $25.00 per year for members; $40.00 for nonmembers, domestic, and $46.00, foreign. Single copies: current issue $10.00 domestic, $11.50 foreign; back issues $11.00 domestic, $12.50 foreign. Circulation 4000. Address communications to:

Vincent J. Tarascio, Managing Editor
Southern Economic Journal
300 Hanes Hall, 019A
Chapel Hill, N.C. 27514

A $15.00 submission fee for SEA members and $25.00 for nonmembers must accompany all manuscripts except book reviews. Membership dues in the Association are: annual $25.00, contributing $42.50, life $325.00, institutions $325.00, student $12.50, retired $12.50. Membership dues include payment for a year's subscription to the *Southern Economic Journal*. Applications for membership should be addressed to Joseph Jadlow, College of Business Administration, Oklahoma State University, Stillwater, Oklahoma 74078.

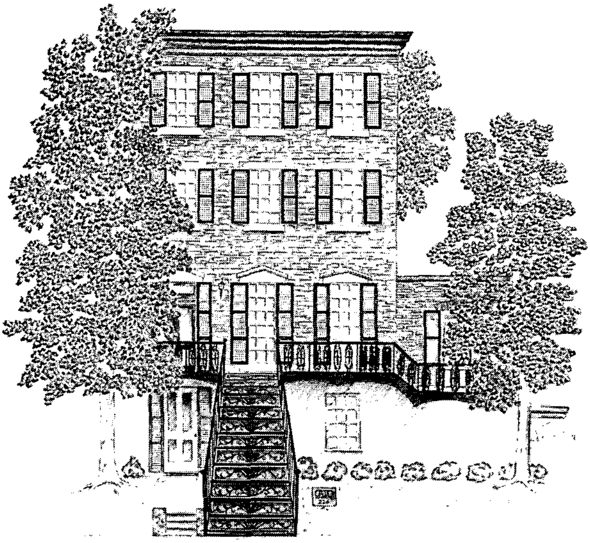
The Cato Institute

Founded in 1977, the Cato Institute is a nonpartisan public policy research foundation headquartered in Washington, D.C. The Institute is named for *Cato's Letters*, libertarian pamphlets that helped lay the philosophical foundation for the American Revolution.

The Cato Institute seeks to broaden the parameters of public policy debate to allow consideration of more options that are consistent with the traditional American principles of limited government, individual liberty, and peace. Toward that goal, the Institute strives to achieve a greater involvement of the intelligent, concerned lay public in questions of policy and the proper role of government.

The Cato Institute undertakes an extensive publications program dealing with the complete spectrum of policy issues. Books, monographs, and shorter studies are commissioned to examine the federal budget, Social Security, monetary policy, natural resource policy, military spending, regulation, NATO, international trade, and many other issues. Major policy conferences are held throughout the year, from which papers are published thrice yearly in the *Cato Journal*. The Institute maintains an informal joint publishing arrangement with the Johns Hopkins University Press.

In order to maintain an independent posture, the Cato Institute accepts no government funding. Contributions are received from foundations, corporations, and individuals, and from the sale of publications. The Institute is a nonprofit, tax-exempt educational foundation under Section 501(c)3 of the Internal Revenue Code.



The Watterston House
headquarters of the
Cato Institute

$6.00

POVERTY AND THE LEGAL FRAMEWORK

A government's social policy helps set the rules of the game—the stakes, the risks, the payoffs, the tradeoffs, and the strategies for making a living, raising a family, having fun, defining what "winning" and "success" mean. The more vulnerable a population and the fewer its independent resources, the more decisive is the effect of the rules imposed from above. The most compelling explanation for the marked shift in the fortunes of the poor [in the 1960s and thereafter] is that they continued to respond, as they always had, to the world as they found it, but that we—meaning the not-poor and un-disadvantaged—had changed the rules of their world. . . . The first effect of the new rules was to make it profitable for the poor to behave in the short term in ways that were destructive in the long term. Their second effect was to mask these long-term losses—to subsidize irretrievable mistakes. We tried to provide more for the poor and produced more poor instead. We tried to remove the barriers to escape from poverty, and inadvertently built a trap.

—Charles Murray

Losing Ground: American
Social Policy, 1950–1980
(New York: Basic Books,
1984), p. 9.